

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:10:41 ; Search time 41 Seconds  
(without alignments)  
502.921 Million cell updates/sec

Title: US-09-869-334b-1  
Perfect score: 2052  
Sequence: 1 MNVLNRQALQRLNGKNK.....SITPIENSVIYGLKSPRYVM 396

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2052	100.0	396	1 YJIB_BACSU	O34374 bacillus su
2	839	40.9	405	1 CPXM_BACSU	P27632 bacillus su
3	738	36.0	410	1 CPXM_BACME	O06069 bacillus me
4	711.5	34.7	410	1 CPXI_BACME	P14762 bacillus me
5	658	32.1	410	1 CPXW_BACSU	O08469 bacillus su
6	646.5	31.5	405	1 CPXW_SACER	P33271 saccharopol
7	600	29.2	395	1 BHOI_BACSU	P53554 bacillus su
8	571	27.8	397	1 CPXQ_SACER	P48635 saccharopol
9	557	27.1	404	1 CPXJ_SACER	O00441 saccharopol
10	548.5	26.7	368	1 CPXW_SULSO	O55080 sulfolobus
11	520	25.3	428	1 C124_MYCTU	O50696 mycobacteri
12	516	25.1	402	1 C123_MYCTU	P77902 mycobacteri
13	514.5	25.1	398	1 C142_MYCTU	O53563 mycobacteri
14	499	24.3	433	1 C125_MYCTU	P71856 mycobacteri
15	489.5	23.9	402	1 C13B_XYLFA	O9965 xylolla fas
16	488	23.8	414	1 C126_MYCTU	P77903 mycobacteri
17	486.5	23.7	404	1 C13B_XYLFT	O87ax5 xylolla fas
18	482	23.5	406	1 CPXO_PSEPU	O59723 pseudomonas
19	475	23.1	405	1 CPXG_STRGO	P18326 streptomyc
20	474	23.1	405	1 C13C_XYLFA	O9966 xylolla fas
21	471.5	23.0	438	1 C140_MYCTU	O08464 mycobacteri
22	465	22.7	399	1 C13C_XYLFT	O87av9 xylolla fas
23	462	22.5	402	1 CPXF_STRGO	P18327 streptomyc
24	451	22.0	405	1 C130_MYCTU	Q11062 mycobacteri
25	446	21.7	381	1 CPXG_STRGO	P23296 streptomyc
26	439	21.4	409	1 CPXJ_STRGO	O59831 streptomyc
27	437.5	21.3	402	1 NOR_FUSOX	P23295 fusarium ox
28	435	20.7	399	1 PASI_RHOFA	P46373 rhodococcus
29	424	20.7	412	1 CPXW_STRGO	P26811 streptomyc
30	419.5	20.4	428	1 CPXJ_PSESP	P33006 pseudomonas
31	418.5	20.4	405	1 CPXW_BACSU	O34926 bacillus su
32	418.5	20.4	406	1 CSC4_AMTOR	O8rn03 amycolatops
33	415	20.2	408	1 NOR2_CYLTO	Q12599 cylindrocac

34	403	19.6	403	1	NOR1_CYLTO	Q00616 cylindrocac
35	401.5	19.6	398	1	C5B3_AMTOR	O8rn04 amycolatops
36	397	19.3	400	1	CPXP_RHISN	P55544 rhizobium s
37	391	19.1	401	1	CPXP_BRAJA	O59203 bradyrhizob
38	390	19.0	436	1	THCB_RHOBR	P43492 rhodococcus
39	385	18.8	489	1	C128_MYCTU	O59572 mycobacteri
40	376	18.3	400	1	C141_MYCTU	O08362 mycobacteri
41	375.5	18.3	429	1	CPXR_BRAJA	O59204 bradyrhizob
42	366.5	17.9	414	1	YAVG_RHISN	O53215 rhizobium s
43	362	17.6	414	1	CPXA_PSEPU	P00183 pseudomonas
44	355	17.3	391	1	C5A3_AMTOR	O8rn05 amycolatops
45	351	17.1	422	1	CPXC_AGRTO	P24466 agrobacteri

## ALIGNMENTS

RESULT 1  
YJIB\_BACSU  
ID YJIB\_BACSU STANDARD; PRT; 396 AA.  
AC O34374  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative cytochrome P450 YJIB (EC 1.14.-.-).  
GN YJIB OR BSU12210.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID:1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RT Rivoita C., Soldo B., Lazarevic V., Joris B., Maue C., Karamata D.;  
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a  
RT putative 12.3 kb operon involved in hexuronate catabolism and a  
RT perfect catabolite-responsive element."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruchli C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot P., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., N.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guilseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogiwara K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic C., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis."  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AF015825; AAC46317.1; --.  
 EMBL; Z99110; CAB13078.1; --.  
 PIR; B69851; B69851.  
 HSSP; Q5080; IIO9.  
 Subtilisin; BG13195; YJ1B.  
 InterPro; IPR001128; Cytochrome\_P450.  
 Pfam; PF00067; P450; 2.  
 PRINTS; PR00385; P450.  
 PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 Complete proteome.  
 FT METAL 349 349 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SO SEQUENCE 396 AA: 44990 MW: 9A899CF13613DBCFF CRC64;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:12:54 ; Search time 3275 Seconds  
(without alignments)  
15762.306 Million cell updates/sec

Title: US-09-869-334B-2  
Perfect score: 1191  
Sequence: 1 atgaatgtgtaaacgcgcg.....gcttcggtgaaatgtaa 1191

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hcg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_scs.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_scs.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1189.4	99.9	35739	1	AF015825	AF015825 Bacillus
2	1189.4	99.9	201241	1	BSUB00007	Z99110 Bacillus su
3	425	35.7	966	6	AX433775	AX433775 Sequence
4	242	20.3	1213	6	AX433767	AX433767 Sequence
5	234.4	19.7	290685	1	AE017032	AE017032 Bacillus
6	229.8	19.3	3187	1	BACRTP	M24523 B.subtilis
7	182.6	15.3	4317	1	BMCTP450A	Z21972 B.megateriu
8	169	14.2	1902	1	BMCP450A	X16610 Bacillus me
9	165.8	13.9	6629	1	AB072568	AB072568 Streptomy
10	162.4	13.6	5008	1	BSUCYPAZL	Y11043 Bacillus su
11	162.4	13.6	17828	1	BSUCYPAZL	Y11043 Bacillus su
12	162.4	13.6	19173	1	BSUB0014	Z99117 Bacillus su
13	159.6	13.4	300893	1	AE017006	AE017006 Bacillus
14	156.4	13.1	10167	1	SFU08223	U08223 Streptomyce
15	150	12.6	342300	1	MLEPRTN8	AL583924 Mycobacte
16	149	12.5	1233	6	AX697983	AX697983 Sequence
17	149	12.5	60196	6	AX697977	AX697977 Sequence
18	145.8	12.2	17512	1	AB071405	AB071405 Lechevali
19	145.8	12.2	25681	1	SAE414559	AJ414559 Saccharot
20	145.8	12.2	26144	1	AB090952	AB090952 Lechevali
21	145.8	12.2	28654	1	AF534707	AF534707 Lechevali
22	144.2	12.1	3078	1	SERCP450A	M83110 Saccharopol
23	144	12.1	39228	1	MLCB1788	AL008609 Mycobacte
24	139.2	11.7	8478	6	AR173058	AR173058 Sequence
25	139.2	11.7	10153	1	BSU51868	U51868 Bacillus su
26	139.2	11.7	19922	1	BSUB0016	Z99119 Bacillus su
27	139.2	11.7	220060	1	AF008220	AF008220 Bacillus
28	138.2	11.6	207829	1	BSUB0010	Z99113 Bacillus su
29	136.4	11.5	6739	1	AB088066	AB088066 Bacillus
30	135.4	11.4	299175	1	AP005023	AP005023 Streptomy
31	125	10.5	300350	1	AP006574	AP006574 Gloebact
32	124.2	10.4	145911	1	AP003014	AP003014 Mesorhizo
33	122.4	10.3	309050	1	SCO939117	AL939117 Streptomy
34	120.6	10.1	10417	1	AE011961	AE011961 Xanthomon
35	120.6	10.1	300885	1	AE012559	AE012559 Xylella f
36	119.2	10.0	10029	1	AE009087	AE009087 Agrobacte
37	119.2	10.0	10184	1	AE008053	AE008053 Agrobacte
38	118.4	9.9	154620	2	EX323047	EX323047 Danio rer
39	118.2	9.9	16057	1	AE007165	AE007165 Mycobacte
40	118.2	9.9	349606	15	EX842583	EX842583 Mycobacte
41	117.8	9.9	24268	1	AP0070947	AP0070947 Streptomy
42	117.8	9.9	300550	1	AP005030	AP005030 Streptomy
43	117	9.8	8407	1	SEU82823	U82823 Saccharopol
44	116.4	9.8	18109	1	AE006571	AE006571 Mycobacte
45	116.4	9.8	320050	1	EX248336	EX248336 Mycobacte

ALIGNMENTS

RESULT 1  
AF015825  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AF015825  
Bacillus subtilis 168 cotT-rapA region sequence.  
35739 bp DNA linear BCT 07-MAY-1998

AF015825  
Bacillus subtilis  
GI:2612880

Bacillus subtilis  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
1 (bases 1 to 35739)  
Rivolta, C., Soldo, B., Lazarevic, V., Joris, B., Mael, C. and Karamata, D.

A 35.7 kb DNA fragment from the Bacillus subtilis chromosome containing a putative 12.3 kb operon involved in hexuronate

Wed



```

/translation="MAGKKTITINGVEMASEEQVLQLLNNSSIEVPOVCVHPSLGP
IETDCTCTVINGELKRSASLKDGDVIDTSLSPDKKQVYIGMDKILYNHELICTVC
DYNNGCGHINTKEMKINHOSIPDPHKPYHKDSEHPFYRDPDQDILLCGRVCAQD
VQVETLTDIWERKEPRVWDNDVPIINSSCVSCCHSTVPCPNAMKEGMEGAYL
TGNNETLRPMLEIKTGVETGYSILAIDMESAMDERIKTKTCTVCTYCGVGSFV
WTKGRDILKVEPQEPAPANGISTCVKGFQWDFVNSEERLTKPLIRGDHFRABWEE
ALLILASFTLEKAFGPDLSUAFITSSKCTNEESYLMQKLRGVIGTNNVNCRSYQ
SPATAGLFRTVGSDGSDITDIAQADLVLIIGNSHESHPVLSTRIKRAHKLGRQKV
IVADIRKHEMARESDLFVQPRAGSDIVMLNAIAKYLIEGKADERFLRVRNGRDEYV
KSLAPYLTLEYAEKRTGIDQETGLIOMAEIMIGQADSDCALWAMGVTOHIGSDTSTAIN
LLVVTGNTGKPGAGSVPLRGHNNVQASDFGSDPDLPCYEKVTDEQVROKYSERVWG
PLPKEPGNTNEMIKIHSQOLKAMYVKEGEMGLVDSNINHVHAAEYKLDFFVQDIF
LSRTAEFADVLVPASLEKGTINTERRIQRLQVFEPLGESKPDQIIMEVANKL
GAGWLYEHPADIMEBAALKSPIYAGVTYERLEGYNLQWPNADKQSPLLFTFRFF
PDGKALYVQWMTPEKFECEYDIHVNNGRLLEHFEHGNLYKSGIKSEKTEPEVLEI
SPELAABERGIDQTLVRLTSPFGNVKVKCLITDRVKGEVLPMDNSGEAAILNLITGS
HADKDTDPAYKETSAKMELIKHDGSLPLKINHNGNPQIQIGVQVHKWKARDYIF
PGDAVKGWGHNG"
7762..8373
/gene="yjd"
RBS
7762..7769
/gene="yjd"
7780..8340
/gene="yjd"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46313.1"
/db_xref="GI:2612889"
translation="WAKAKRIOKLEVTEDORKEDELEIDALIDHKEAILETHML
GHMNERGVLPRLRGLFGQDGKVLIDLVKKADEETPANTLKNLLIFGLMDLVKQLE
FLILKNAVAGASAVEKQNSEKTYGFDIIRSLKDPPEINKSLITLPSFLKMGQDPTKEL
ERTTQPPEHQHQBPREKRNKRD"
8347..8373
/gene="yjd"
RBS
8525..9223
/gene="yja"
8525..8530
/gene="yja"
8537..9178
/gene="yja"

```

```

Query Match      99.9%; Score 1189.4; DB 1; Length 35739;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAATGTGTTAAACCGCGGCAAGCCTTGACGAGCGCTGCTCAATGGGAAAAACAAA 60
DB 10583 ATGAATGTGTTAAACCGCGGCAAGCCTTGACGAGCGCTGCTCAATGGGAAAAACAAA 10642

QY 61 CAGGATGCGTATCATCCGTTTCCATGCTGATGAATCGATGAGAAAGATCGCGCTTTTCC 120
DB 10643 CAGGATGCGTATCATCCGTTTCCATGCTGATGAATCGATGAGAAAGATCGCGCTTTTCC 10702

QY 121 TTTGATGAGAAAACCAAGTGTGGACGCTTTTCTTTATGATGATGTCAAAAAGTTGTT 180
DB 10703 TTTGATGAGAAAACCAAGTGTGGACGCTTTTCTTTATGATGATGTCAAAAAGTTGTT 10762

QY 181 GGGGATAAAGAGTGTGTTTCCAGTTTGCATGCGGACGACGACGCTCTATTGGAATTTCC 240
DB 10763 GGGGATAAAGAGTGTGTTTCCAGTTTGCATGCGGACGACGACGCTCTATTGGAATTTCC 10822

QY 241 ATCAATTAACATGACCGCGGAGCAGATACAAAATCCCGTTTCAGTGTGAGCAAAAGCCTTT 300
DB 10823 ATCAATTAACATGACCGCGGAGCAGATACAAAATCCCGTTTCAGTGTGAGCAAAAGCCTTT 10882

QY 301 ACTCCGCGGTGATGAAGCAATGGGACCGGAAATTCAGAAATCACAGATGAATGATT 360
DB 10883 ACTCCGCGGTGATGAAGCAATGGGACCGGAAATTCAGAAATCACAGATGAATGATT 10942

QY 361 CAAAATTTCCAGGCGGACGATGAGTTGACCTTGTTCAGATTTTTCATACCGCTTCG 420
DB 10943 CAAAATTTCCAGGCGGACGATGAGTTGACCTTGTTCAGATTTTTCATACCGCTTCG 11002

```

```

QY 421 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCA 480
DB 11003 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCA 11062

QY 481 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAAGTGAAGAGCTGGAAGAGCCTTT 540
DB 11063 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAAGTGAAGAGCTGGAAGAGCCTTT 11122

QY 541 TTGGAAGAACGAGATAACTGTGAGGAAGAACTGGCCCGCTTTTTCGCCGGCATCATAGAA 600
DB 11123 TTGGAAGAACGAGATAACTGTGAGGAAGAACTGGCCCGCTTTTTCGCCGGCATCATAGAA 11182

QY 601 GAAAGCGGAAACAAACCGGAAACAGGATATTATTTCTATTATTAGTGAAGCGGAAAGAAACA 660
DB 11183 GAAAGCGGAAACAAACCGGAAACAGGATATTATTTCTATTATTAGTGAAGCGGAAAGAAACA 11242

QY 661 GCGGAGAAAGCTGTGCGGTGAAGAGCTGATTCGTTTTTGACGCTGCTGCTGTGTCGCCGGA 720
DB 11243 GCGGAGAAAGCTGTGCGGTGAAGAGCTGATTCGTTTTTGACGCTGCTGCTGTGTCGCCGGA 11302

QY 721 AATGAACCACTACAAACCTGATTTCAATCGATGTACAGCATATTAGAACGCGCAGC 780
DB 11303 AATGAACCACTACAAACCTGATTTCAATCGATGTACAGCATATTAGAACGCGCAGC 11362

QY 781 GTTTACGAGGAACTGCGCAGCCATCTCTGAACCTGATGCTCAGGCAAGTGGAGGAAAGCCTTTG 840
DB 11363 GTTTACGAGGAACTGCGCAGCCATCTCTGAACCTGATGCTCAGGCAAGTGGAGGAAAGCCTTTG 11422

QY 841 CGTTTCAGAGCGCGCGCGCGCGCTTTGAGGGCGCATTCGCAAGCGGATACGAGATCGG 900
DB 11423 CGTTTCAGAGCGCGCGCGCGCGCTTTGAGGGCGCATTCGCAAGCGGATACGAGATCGG 11482

QY 901 GGGCACCTGATTAAAGAAAGTGTATGTTTTGGGCTTTTGGCATCGGCAATCGTGAT 960
DB 11483 GGGCACCTGATTAAAGAAAGTGTATGTTTTGGGCTTTTGGCATCGGCAATCGTGAT 11542

QY 961 GAAGCAAAAGTTTGACAGCCGACATGTTTGATATCGCGCGCATCCCAATCCGCATATT 1020
DB 11543 GAAGCAAAAGTTTGACAGCCGACATGTTTGATATCGCGCGCATCCCAATCCGCATATT 11602

QY 1021 GCGTTGCGCAGGATCCATTTTTCCTTGGGCGCGCTGCGCGCTTTCGAGCAAT 1080
DB 11603 GCGTTGCGCAGGATCCATTTTTCCTTGGGCGCGCTGCGCGCTTTCGAGCAAT 11662

QY 1081 ATCGGTTAAAGCTCTTTGATTCTGCTTTCTCATATGAGTGTGCTCAGTATCACTCG 1140
DB 11663 ATCGGTTAAAGCTCTTTGATTCTGCTTTCTCATATGAGTGTGCTCAGTATCACTCG 11722

QY 1141 ATTGAACACGTGTGATATACGATTAAAGAGCTTTCGTTGAAAATGTAA 1191
DB 11723 ATTGAACACGTGTGATATACGATTAAAGAGCTTTCGTTGAAAATGTAA 11773

```

```

RESULT 2
BSUB0007      BSUB0007      201241 bp      DNA      linear      BCT 07-JUL-2003
LOCUS         Bacillus subtilis complete genome (section 7 of 21): from 1209742
DEFINITION    to 1410982.
ACCESSION     Z99110 AL009126
VERSION       Z99110.2 GI:32468738
KEYWORDS      Bacillus subtilis subsp. subtilis str. 168
SOURCE         Bacillus subtilis subsp. subtilis str. 168
ORGANISM      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE     1 (bases 1 to 201241)
AUTHORS      Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrati, E.,

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:01:12 ; Search time 223 Seconds  
(without alignments)  
560.292 Million cell updates/sec

Title: US-09-869-334B-1  
Perfect score: 2052  
Sequence: 1 MNVLRQALQALLGNK.....SITPIENSVIYGLKSRVNM 396

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_protist.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rviro.\*
- 16: sp\_bacterioph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	833.5	40.6	404	16	Q81Q07	Q81Q07 bacillus an
2	732	35.7	409	16	Q81CY0	Q81CY0 bacillus ce
3	715	34.8	411	16	Q81Q12	Q81Q12 bacillus an
4	674	32.8	417	2	Q59910	Q59910 streptomyce
5	661.5	32.2	402	2	Q8R221	Q8R221 streptomyce
6	642	31.3	416	2	Q87605	Q87605 streptomyce
7	634.5	30.9	393	16	Q82LM3	Q82LM3 streptomyce
8	626	30.5	376	16	Q81785	Q81785 bacillus su
9	622	30.3	457	16	Q82BH2	Q82BH2 streptomyce
10	601	29.3	395	2	Q8K2M6	Q8K2M6 bacillus su
11	599.5	29.2	388	16	Q82A10	Q82A10 streptomyce
12	589	28.7	433	16	Q9F2Q0	Q9F2Q0 streptomyce
13	582.5	28.4	397	2	Q8KIC8	Q8KIC8 nocardia ae
14	581	28.3	397	2	Q59523	Q59523 micromonospor
15	581	28.3	434	16	Q9CBE7	Q9CBE7 mycobacteri
16	576.5	28.1	430	16	Q8NSW2	Q8NSW2 corynebacte

17	575.5	28.0	406	2	Q83X78	Q83X78 streptomyce
18	571	27.9	418	16	Q9HIR4	Q9HIR4 pseudomonas
19	570.5	27.8	405	16	Q82ES4	Q82ES4 streptomyce
20	566	27.6	401	2	Q9KHJ7	Q9KHJ7 streptomyce
21	565	27.5	399	2	Q8GGQ1	Q8GGQ1 streptomyce
22	564	27.5	404	2	Q54302	Q54302 streptomyce
23	562.5	27.4	413	2	Q848C0	Q848C0 streptomyce
24	561.5	27.4	417	2	Q83WG3	Q83WG3 streptomyce
25	560.5	27.3	433	16	Q89G11	Q89G11 bradyrhizob
26	557.5	27.2	410	2	Q9XSP9	Q9XSP9 streptomyce
27	550	26.8	399	16	Q82QC6	Q82QC6 streptomyce
28	548.5	26.7	404	2	Q83X67	Q83X67 corynebacte
29	547.5	26.7	461	16	Q8FS42	Q8FS42 streptomyce
30	545	26.6	419	2	Q8X332	Q8X332 streptomyce
31	544.5	26.5	407	2	Q59819	Q59819 streptomyce
32	540	26.3	399	16	Q93HA3	Q93HA3 mycobacteri
33	538	26.2	310	2	Q32927	Q32927 mycobacteri
34	536.5	26.1	411	16	Q9X8Q3	Q9X8Q3 streptomyce
35	534.5	26.0	367	17	Q97212	Q97212 sulfolobus
36	533	26.0	396	2	Q9L465	Q9L465 streptomyce
37	532.5	26.0	411	2	Q32460	Q32460 actinomadu
38	532	25.9	404	16	Q93HJ0	Q93HJ0 streptomyce
39	531.5	25.9	418	16	Q89NK6	Q89NK6 bradyrhizob
40	529.5	25.8	417	2	Q8R2Z4	Q8R2Z4 stigmatella
41	528	25.7	415	16	Q92QQ5	Q92QQ5 rhizobium m
42	527	25.7	396	2	Q9EYL2	Q9EYL2 streptomyce
43	524	25.5	394	16	Q9F2L1	Q9F2L1 streptomyce
44	524	25.5	415	16	Q8UFY8	Q8UFY8 agrobacteri
45	523.5	25.5	416	16	Q982V9	Q982V9 rhizobium l

## ALIGNMENTS

## RESULT 1

Q81Q07 PRELIMINARY; 2RT; 404 AA.  
ID Q81Q07  
AC Q81Q07;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Cytochrome P450 family protein.  
GN BAZ632.  
OS Bacillus anthracis (strain Ames).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=198094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608414; PubMed=12721629;  
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,  
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
RA Benton J.B., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
RA Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria.";  
RL Nature 423:81-86(2003).  
DR EMBL; AE017832; AAP26430.1; -  
DR TIGR; BAZ632; -  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; p450; 1.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Complete proteome.  
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;

Query Match

40.6%; Score 833.5; DB 16; Length 404;

```

QY 14 LLGNKNQDAYHPFZMYESMRKDPVSGDEENQV---MSVFYDDVKKVGVGKEJ----- 65
| : : ||| : | | | : : : | : : | : |
Db 17 LASAQFEDAYE---YKESRKQVPLFVNKTELGAEMLITRYEDALPZCKONRLKKCPA 73
QY 66 ---FSSCMPQOQTSIGNS-----IINMPPKHTKIRVVVNKEZFPRVMQWEPRIQEITD 117

```

Best Local Similarity 39.2%; Pred. No. 8.5e-44;

Matches 148; Conservative 78; Mismatches 126; Indels 2

14 \* I NCNINW00XVHDEBWE CMBVT8 DUCENENOW -- WSEVEI YDNINIKXZAEIC

14	LLGNNAQDAI:FFIEMIESRKTAPVSFDEENQV	--NSVTFIDDDVNAV
	: : :           : : :   : :	==
	: : :           : : :   : :	==

17 LASAQFKEDAYE---YKESRKVQPVLFNKTELGAEWLITRYEDALPLEK

Y  
66 --FSSCMPQQTSSIGNS-----IINMDPPKHTKIRSVVNKAETPRVMKQW

```

Db 74 NVFSODTLNVLFTVNSDYLTHMLNSOPPHNRLRSIVQVFTPKMIAQLGRIQDIAD 133
Qy 118 ELIOKQFQSRFSDLVHDSYPLVPIVIGELGVPSAHMEQKAWGDLVSTPKRSEAE 177
Db 134 DLLNEVERKSGSLNVDDYSFPLPIIVISEMIGIPKEDQAKRWIMSHAVIRPETEE--- 190
Qy 178 KAFLEERDKCEELAAFFAGIIEKRNKPEODIISILVEAETGEKLSGEEILIPFCTLL 237
Db 191 ---IKETEKQLSEFTYIQLVDMKKEPKEDVLSALILAESEGHKSARELYSMIMLLI 247
Qy 238 VAGNETTNLISNAMYSLIETPGVVEERSHPELMPQAVEALRFPAPV-LRRIAKRD 296
Db 248 VAGHETTNNLTNTVALLLENQQLKENPKLIDAAIEGLRYSPVEVTTSMWADPE 307
Qy 297 TEIGHLKEGDMVLAFAVANRDKAKDRPMFDIRHPNPHIAFGHICGLAPLAR 356
Db 308 FOIHQOTIEKGMVWVIALAANRDETVEFENPEVFDITRENNRHIAFGHSGHFCGLAPLAR 367
Qy 357 LEANIALTSLISAFPHME 374
Db 368 LEAKIATTLFRMPELQ 385

RESULT 4
ID Q59910 PRELIMINARY; PRT; 417 AA.
AC Q59910;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative cytochrome P450 113B1.
GN CYP113B1.
OS Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T59235;
RX MEDLINE=95075319; PubMed=7984112;
RA Merson-Davies L.A., Cundliffe E.;
RT "Analysis of five tylosin biosynthetic genes from the tyll3A region of
RT the Streptomyces fradiae genome."
RL Mol. Microbiol. 13:349-355(1994).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U08223; AAA21341.1; -.
DR FIR; S49051; S49051.
DR HSGP; Q55080; I107.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Electron transport; Heme; Membrane;
KW Monooxygenase; Oxidoreductase.
FT METAL 366
SQ SEQUENCE 417 AA; 47224 MW; A761942242PC35B1 CRC64;

Query Match 32.8%; Score 674; DB 2; Length 417;
Best Local Similarity 40.2%; Pred. No. 8.3e-41;
Matches 144; Conservative 64; Mismatches 126; Indels 24; Gaps 4;

Qy 29 WYEMRKDAPVSPDEENQVMSVFLYDDVKVVGDKELFSS-----CMPQCTSSI--GNS 80
Db 38 WFAARAEAPVFWDESQRQAWQVFRYDDYTLVTNPNQFSSDFSPVPVPEELALMGPGT 97
Qy 81 IINMPDPKHTKRSVWNKAPTRVMKQWBPRIQETDELTKQFGRSEFDLVHDFSYPLP 140
Db 98 FCGIDPPRNGPLKLVSAQFTPRATLPRIAETIGLLDGLRKGQIDVDSLDLAYPLP 157
Qy 141 VIVISLGVPSAHMEQKAWGDLVST-----PKKSEAEAKAFLEERDKCEE 192

```

```

Db 158 VVIAELGIPABDRDLFRWVDVVIKNGMEYPNLPDDFSETMGPAIKEMGSDYLFRRA 217
Qy 193 AFFAGIIEBKRNKPEODIISILVEAETGEKLSGEEILIPFCTLLVAGNETTNLISNAM 252
Db 218 L-----KRETPDDLSMGLIEAEVEGRRLLTDEIVNIVALLTAGHSSATLLGNLF 269
Qy 253 YSILSTPGVVEERSHPELMPQAVEALRFPAPVLRRIAKRDTEIGHLKEGDMVLA 312
Db 270 LVLDHREBAQELRADROLIPGAIBETLRYSPFNIFELKEDTDILGHPMKAGOMVVA 329
Qy 313 FVASANRDEAKDRPMFDIRHPNPHIAFGHICGLAPLARLEANIALTSLISAF 370
Db 330 WIASANRUSAHFSDPTDFVRQPNKHMFGHICGLSFLARLEAKVLELFFDFEF 387

RESULT 5
ID Q8RR21 PRELIMINARY; PRT; 402 AA.
AC Q8RR21;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 monooxygenase.
GN VISO.
OS Streptomyces virginiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1961;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21940635; PubMed=11943483;
RA Namwat W., Kamioka Y., Kinoshita H., Yamada Y., Nihira T.;
RT "Characterization of virginiamycin S biosynthetic genes from
RT Streptomyces virginiae."
RL Gene 285:283-290(2002).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB072568; BAB83674.1; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PRO0385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 402 AA; 44632 MW; 622ADF33CFB1C946 CRC64;

Query Match 32.2%; Score 561.5; DB 2; Length 402;
Best Local Similarity 40.1%; Pred. No. 6.4e-40;
Matches 146; Conservative 69; Mismatches 130; Indels 19; Gaps 7;

Qy 29 WYEMRKDAPVSPDEENQVMSVFLYDDVKVVGDKELFSSCMPQ---QTSIGNKIN-M 84
Db 22 WLRTMEHPVHEDYGV-VFHVYHSDVLAVTSDPAVFSSDLRLRFDSSALSEEILSVI 80
Qy 85 DPPKHTKRSVWNKAPTRVMKQWBPRIQETDELTKQFGRSEFDLVHDFSYPLVIVI 144
Db 81 DPPLHKLRLSVSAFTLRVADLEPRVTELAGRLLEKVEG-SEFOLWGFAYPLVIVI 139
Qy 145 SELLGVPSAHMEQKAWGDLVSTPKDSEAKAFLEERDKCEE-----ELAAFFA 196
Db 140 AELLGVPAEDRELFRGWSDRMLSMQVD--DPLIOFGDEAGEDYERLVKPLKEXHYLQ 197
Qy 197 GIIEEKRNKPEODIISILVEAETGEKLSGEEILIPFCTLLVAGNETTNLISNAXYSIL 256
Db 198 RHVDARETPGDDLSLVTAEIAGERLTDQIVFEGALLMAGHYVSTSLGKTVLCLE 257
Qy 257 ETGVTVEERSHPELMPQAVEALRFPAPVLRRIAKRDTEIGHLKEGDMVLAFAVAS 316
Db 258 ENDETAALRADRALISGVIEEVLRMRPPIITVAARVTTGEVWVGVTIPKDRVMVASLLS 317
Qy 317 ANDEKAKEPDMFDRIRHPNPHIAFGHICGLAPLARLEANIALTSLISAFPHMECV 376

```

```

Db 318 ANHDERHIQDPEVDFRRSPNPQLAFGHGHIYCLGSPGLARLEGRVALEMLLDRF---EDI 374
QY 377 SITP 360
Db 375 RVTP 378

RESULT 6
ID O87605 PRELIMINARY; PRT; 416 AA.
AC O87605;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 CYP107L1.
GN CYP107L1 OR PICK OR PIKC.
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1936;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC15439;
RC STRAIN=ATCC15439;
RX MEDLINE=99051447; PubMed=9831532;
RA Xue Y., Wilson D., Zhao L., Liu H.-W., Sherman D.H.;
RT "Hydroxylation of macrolactones VC-17 and narbomycin is mediated by
RL the p1KC-encoded cytochrome P450 in Streptomyces venezuelae.";
RN Chem. Biol. 5:661-669 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15439;
RX MEDLINE=98445333; PubMed=9770448;
RA Xue Y., Zhao L., Liu H.W., Sherman D.H.;
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces
RL venezuelae: architecture of metabolic diversity.";
RN Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116 (1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF087022; AAC64105.1; -.
DR EMBL; AF079139; AAC68886.1; -.
DR HSSP; Q00441; IOXA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 416 AA; 46038 MW; B7392C742045F06B CRC64;

Query Match 31.3%; Score 642; DB 2; Length 416;
Best Local Similarity 37.4%; Pred. No. 1.7e-38;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AHPFPWYENRKPDP---VSFDENQVSVFLYDDVKKVGD---KELFSSCMP- 71
Db 22 QDFAADPYPTAKLRAGSPAHVRTPGDEVLVGVYDRAVLADPRFSKDWENSTTFL 81
QY 72 -QQTSSIGNSTINMDPKHTKIRSVNKAFTPRVMQWEPRIQETIDELQKF---QGR 126
Db 82 TEAEALNNHNLMDSPRRHTRKRLKVLRAEFTRRVLLRPVQVETDGLVDAMLAADGR 142
QY 127 SEFDLVHDFSVPLVIVISLLGVPSAHMEQFKAWSLLVSTPKDKSEAEKAFLEERDK 186
Db 142 A--DLMESLAWPLDITWISLLGVPEPDRAAFRVMTDAFVP--PDDPAQAQTAM----- 191
QY 187 CESELAAPFAGIIEKRNKPEQDIISIVE-ABETGEXLSEELIPFCTLLLVAGNETTT 245

```

```

Db 192 --AEMSGYSLRLLDSKRGQGDGELLSSALVTSDEDSRLTSEELGMAHILLVAGHETTV 249
QY 246 NLISNAMYSILETPGVYEEELRSHPELMPQAVEALRFRAPA-PVLRRIAKEDTEIGGELI 304
Db 250 NLIANGYALLSHPDQLAALRADMTLLDGAVESMLRYEGPVESATYRFPVPVLDGTVI 309
QY 305 KEGDMVLAFAVANRDEAKFDPHMFDIRHPNPHIAFGHGHICGAPLARLEANTALT 364
Db 310 PAGDTVLVLADAHRTPEPFPDRFDIRDTAGHLAFGHGHIHFCIGAPLARLEARIAR 369
QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSFVK 395
Db 370 ALLERCPDL-ALDVSPGELVWYVNPMPIRGLKALFIR 404

RESULT 7
ID Q82LM3 PRELIMINARY; PRT; 393 AA.
AC Q82LM3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative cytochrome P450.
GN CYP8 OR SAV1987.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=12692562;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RL avermitilis: deducing the ability of producing secondary
RL metabolites.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RL microorganism Streptomyces avermitilis.";
RN Nat. Biotechnol. 21:528-531 (2003).
DR EMBL; AP005029; BAC69698.1; -.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Complete proteome.
SQ SEQUENCE 393 AA; 43126 MW; 1573C19D9A256BA3 CRC64;

Query Match 30.9%; Score 634.5; DB 16; Length 393;
Best Local Similarity 40.2%; Pred. No. 5.7e-38;
Matches 145; Conservative 61; Mismatches 132; Indels 23; Gaps 6;

QY 26 PFPYTESMRKDAPVSF-----DEENQVSVFLYDDVKKVGDKELFSSCYQQTSSI--- 77
Db 18 PYPVYALRERGPVHVTRTPPEAFEGVLVGVHEEARAALADPRLSKDGTKKGLTSLDVE 77
QY 78 --GNSIINMDPKHTKIRSVNKAFTPRVMQWEPRIQETIDELQKFQGRSEFDLVHDF 135
Db 78 LMGYLVLDVDPPEHTRLSLVARAFTRVREVALPRIQETIDGLDEMLPRGRADLVDSF 237
QY 136 SYPLPVIVISLLGVPSAHMEQFKAWSLLVSTPKDKSEAEKAFLEERDKCEELAAFF 195
Db 138 APLPVTIVICELGVDPIDIRVTFRLSNIVAPITGGDAELA-----AYERLAAYL 187

```

Qy	196	AGIIEKRN-KPSQDIISILWEAR-ETGEKUSGEELIPFCTILLVAGNETTNLSINAMY	255
Db	188	DELIDDKRSTAPADDLLGDLIRTRAEDDDRSAGEELRAMAFILLVAGHETTNNLITNGVH	247
Qy	254	SILETRGVWEELRSHPELMPQAVREARFRAPAPVLR-RIAKRDTWIEGHLIKEGDMVLA	312
Db	248	TLTHPDLQALRADMTLLDCAVEVLRFEGEVETATTRYAESWEIGSTAIAEGDPVMI	307
Qy	313	FVASANDEAKFRDPHMFDIRRHPNPHIAGFHGIFCLGAPLARLEANIATLSISAPFH	372
Db	308	GLDAAAGRDPAHPDPHVFDIRAPQGHLAGFHGHIYCLGAPLARLEARVALRSILERC	367
Qy	373	M 373	
Db	368	L 368	

RESULT 8

O31785 PRELIMINARY; PRT: 376 AA.

ID O31785 PRELIMINARY; PRT: 376 AA.

AC O31785;

DT 01-JAN-1998 (TReMBUrel. 05, Created)

DT 01-JAN-1998 (TReMBUrel. 05, Last sequence update)

DT 21-JUN-2003 (TReMBUrel. 24, Last annotation update)

DE Putative cytochrome P450 107K1.

GN PKSS OR CYP107K1.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA MEDLINE=38044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriello R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pulic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,

RA Viari A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.";

RT Mature 390:249-256(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; Z99113; CAB13607.1; -

DR PIR; G69679; G69679.

RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RL Nat. Biotechnol. 21:526-531(2003).  
 DR EMBL; AP005035; BAC71248.1; --  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 457 AA; 50512 MW; 8342429F7650A635 CRC64;

Query Match 30.3%; Score 622; DB 16; Length 457;  
 Best Local Similarity 33.3%; Pred. No. 5.6e-37;  
 Matches 137; Conservative 76; Mismatches 148; Indels 50; Gaps 9;  
 QY 23 AYHPPWYEMRKQAPV---SPDEENQVMSVFLYDDVKKVGVGDKELFSSCMPQQTSSIGN 79  
 DB 55 ATDPYFAYAWLREHAPVHRTTLPSPGVAEALVTVYADAKQALADARLSKQPVHHSADPGK 114  
 QY 80 S-----IINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQIETDELICKFQ 124  
 DB 115 SKTGIPGERSANLTHLMDPDHTRLRRLVSKAFTPRVAFAFRVQELTDHLIDQPA 174  
 QY 125 GRSEDLVHDFYPLPVIVISELLGVPSAHMEOFKAWSDLLV---STPKDKSEAEKAPL 181  
 DB 175 QTSADLIHEFAFPLPIYAICDLGVPRDQDDFRDWAGMMIRHGGGPRGVARSVK--- 231  
 QY 182 EERDKCEEELAAFFAGIIEKRNK-----PEODIISILVEAEETGKLSGEELIPFCTL 235  
 DB 232 -----KORGVLAEIIRKREALPADPGGEDLLISGLIRASDHGEHLTENEAACVF 283  
 QY 236 LLVAGNETTNTLISNAMYSLKETPGVYBEL-----RSHPELMPQAVEALRFRAPAPVLR 290  
 DB 284 LLFAGFETTINLIGNTYALLRNPPQARLQSSIERGEQDQLDTGIELLYDGPVELAT 343  
 QY 291 -RIAKKDTIGHLIKEGDWLAFVASANDKAFDRPHMFDIRRHNPNHIAFGHGHFC 349  
 DB 344 WRVATEPLDGGGRIASGDPVILVLAARDPARFDEPTDLDSRSDNQHLYGHHGHC 403  
 QY 350 LGAPLARLEANTALTSLSAFHMECVSTIPIE-----NSVIYGLKSPRYK 395  
 DB 404 IGAPLARLESGRTALTLRLRLPDLR-LAADPADLRWGGGLMRGLNLPVE 453

RESULT 10  
 Q8K2M6 PRELIMINARY; PRT; 395 AA.  
 AC Q8K2M6;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Cytochrome P450 enzyme.  
 GN BIOI.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OK2.  
 RA Sakaki M., Kotanagi T., Kurusu Y.;  
 RT "Genetic analysis of biotin operon in Bacillus subtilis natto OK2";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL; AB088066; BAC0324.1; --  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Heme; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 395 AA; 44869 MW; 6D4B5C445107D225 CRC64;

Query Match 29.3%; Score 601; DB 2; Length 395;  
 Best Local Similarity 34.6%; Pred. No. 1.5e-35;  
 Matches 125; Conservative 85; Mismatches 135; Indels 16; Gaps 6;  
 QY 25 HPPPWYEMRKQAPV---SPDEENQVMSVFLYDDVKKVGVGDKEL-----FSSCMPQQT 74  
 DB 15 NPYSFYDTRAVHPIYKGSF-LKYPGWYTVGHEETAAILKQARFKVTPFEESTKYQDL 73  
 QY 75 SSIGNS-IINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQIETDELICKFQGRSEFLVH 133  
 DB 74 SHVQNMFLFQNPQDHRRLRLTLASGFTPRATESYQPIETVHQLLDQVQGEKQKEVIS 133  
 QY 134 DSYPLPVIVISELLGVPSAHMEOFKAWSDLLVSTPKDKSEAEKAPLEERDKCEEELAA 193  
 DB 134 DFAFPLASVFVIANIIGVPEEDRELQKEWASLIQT---IDFTRSRKVLTEGNNHMAVOAMA 190  
 QY 194 FPAGITTEKRNKPEODIISILVEAEETGKLSGEELIPFCTLVAGNETTNTLISNAMY 253  
 DB 191 YFELLQKRKHHPQOMISMLKQKE-NDKLTEEAASCTILLAIAGHETTVNLISNVL 249  
 QY 254 SILETPGVYEEELSHPELMPQAVEALRFRAPAPVLRRIAKRDTETIGHLIKEGDWVLA 313  
 DB 250 CLIQHPEQLLKRENDPLIGTAVEECLRYESPQMTARVASEDIDISGVTIROGEQVYL 309  
 QY 314 VASANDKAFDRPHMFDIRRHNPNHIAFGHGHFCGAPLARLEANTALTSLSAFPHM 373  
 DB 310 LGAANDRPSIFTNPDVFDITRSPNPLSFHGHHVCLGSLARLEAQIAINTFLQRPESL 369  
 QY 374 E 374  
 DB 370 K 370

## RESULT 11

Q82A10 PRELIMINARY; PRT; 388 AA.  
 ID Q82A10;  
 AC Q82A10;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative cytochrome P450.  
 DE CYP23 OR SAV6249.  
 GN Streptomyces avermitilis.  
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=21477403; PubMed=11572948;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 RT avermitilis: deducing the ability of producing secondary  
 RT metabolites";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=12692562;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinoe M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RL Nat. Biotechnol. 21:526-531(2003).  
 DR EMBL; AP005046; BAC73960.1; --  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 388 AA; 42468 MW; 1327D228BF068E647 CRC64;

Query Match 29.2%; Score 599.5; DB 16; Length 388;  
 Best Local Similarity 34.5%; Pred. No. 1.9e-35;  
 Matches 128; Conservative 73; Mismatches 143; Indels 27; Gaps 5;

QY 26 PFPYVESMRKDAVP---SFDEENQVMSVFLYDDVKKVGVGDKELFSSCMPQQTSSIGN---79  
 DB 11 PYAVYDRLEDTPAVHRIAGTDGKPAWLVTYRDDVREGIANPLL---SLDKKHALPGNYRG 67

QY 80 -----SINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIOKFGQRSEPD 130  
 DB 68 LALPALDANLLNMDAPDHTIRRLVGRAFTLRRVVEQLSEPVRETAHRLDLGTHGSTD 127

QY 131 LVHDFSYPVIVISILGVPSAHMEQKAMSDLLVSTPKQSEBAKAFLEERDKCEE 190  
 DB 128 LIASVAAFLPTIVICDILGVPEHRDFRAWTDLVTPDPARPDVARESVW-----S 179

QY 191 LAAPFAGIIEEKXNPEQDIISILVEAETGEKLSGEBELIPCTILLVAGNETTINLSN 250  
 DB 180 LLGFTGSLADKRKNPADDLLSDLIABVEEGDRLTDELSLAFILFAGYENTVHLIGN 239

QY 251 AMYSILETPGVYEECSRSHPELMPQAVEALPRAPAPV--LRRIAKRDTIEIGGLIKEGDM 309  
 DB 240 AVLALLRHEQOLAA--REDPARLPDVGSEFARVEGPAALLAIREFPVVDVTIGGVTPAGET 299

QY 310 VLAFVASANRDEAKEDRDMEDIRHPNPHIAPGHIHFCIGAPLARLEANIALTSLISA 369  
 DB 300 VLLSLSAANRDSRPFDPDLGRDAAGHALGHGVHYCLGAPLARLETEVALAALLER 359

QY 370 FPHMECVSITP 380  
 DB 360 FPDIALAETEP 370

RESULT 12

Q9F2Q0 PRELIMINARY; PRT; 433 AA.  
 AC Q9F2Q0;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative cytochrome P450 hydroxylase.  
 GN SC03099 OR SCE41.08C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1];  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2];  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3];  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Kinashi H., Hopwood D.A.;  
 RL "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4];  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieger H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S.,

Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT \*Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).;  
 RL Nature 417:141-147(2002).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL; AL939115; CAC09540.1; --  
 DR HSSP; C00441; LOXA.  
 DR GO; GO:0004497; F-monooxygenase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 433 AA; 47351 MW; CDED6D74CA4DA508 CRC64;

Query Match 28.7%; Score 589; DB 16; Length 433;  
 Best Local Similarity 32.7%; Pred. No. 1.3e-34;  
 Matches 134; Conservative 75; Mismatches 151; Indels 50; Gaps 9;

QY 23 AVHPFPEYVESMRKDAVP---SFDEENQVMSVFLYDDVKKVGVGDKELF-----66  
 DB 18 ASDPFAVAMLRKHAHPVHTRLPSPGVAEWLVTRYADAKQALADPRLSKNPAHHDEPAHAK 77

QY 67 --SSCMPQQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIOKFG 124  
 DB 78 GKTGIPGERKAEKELMTHLINIDPDHTLRLRLVSKAFTPRVAEFAFVQELADGLDRFA 137

QY 125 GRSEFDLVHDFSYPVIVISILGVPSAHMEQKAMSDLLV---STPKDKSEAEKAF 181  
 DB 138 DTGSADLIHDFAPLPPIVAICDLGVPRDQDDFRDWAGMMIRHQGGPRGVARSVK---194

QY 182 EERDKCEELAAFFAGIIEEK-----NKPEQDIISILVEAETGEKLSGEBELIPCTL 235  
 DB 195 -----KRGYLAHLIRKRAALPEPAPEGDLISGLIRADSGEHLITENEAAMAFI 246

QY 236 LLVAGNETTINLSNMYSILETGPVYEEIRS-----HPELMPQAVEALRFRAPAPVLR 290  
 DB 247 LLFAGFETTVNLWGNTYALLTHPEQRELRQTSLAAGERGLLETGVLEELRYDGPVELAT 366

QY 291 -RIAKRDTIEIGGLIKEGDMVLAFVASANRDEAKFDRPHMFDRRHPNPHIAPGHIHFC 349  
 DB 307 WFATRPPLTIGGQVAAAGDPVLWLAADRDPERFTDPTDLIARDSQLHGYGHIHC 366

QY 350 LGAPLARLEANIALTSLISAFPHMECVSITPIE-----NSVTYGLKSFV 394  
 DB 367 LGAPLARLEGQTALATLTLRLFDLR-LAADPAELRWGGILNRGLTLPV 415

RESULT 13

Q8KIC8 PRELIMINARY; PRT; 397 AA.  
 AC Q8KIC8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Cytochrome P450 enzyme (Putative cytochrome P450) (Putative P450 protein) (Cytochrome P-450 Rebd).  
 DE RBME OR REBT.  
 OS Nocardia acrocolonigenes.  
 GN Nocardia acrocolonigenes.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Pseudonocardiaceae; Actinosynnemataceae; Lechevalieria.  
 OX NCBI\_TaxID=68170;  
 RN [1];  
 RP SEQUENCE FROM N.A.  
 RA Hyun C.-G., Billiign T., Liao J., Thorson J.S.;  
 RT "The Biosynthesis of Indolocarbazoles in a Heterologous E. coli Host."  
 RL Submitted (AJG-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2];  
 RP SEQUENCE FROM N.A.



RA Nishizawa T., Sherman D.H.;  
RT "Identification and analysis of the rebeccamycin biosynthetic gene  
RT cluster in *Lechevalieria aerocolonigenes* ATCC39243.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 39243;  
RX MEDLINE=21979046; PubMed=11983340;  
RA Sanchez C., Brana A.F., Mendez C., Salas J.A.;  
RT "Rebeccamycin biosynthesis gene cluster.";  
RL Chem. Biol. 9:519-531 (2002).  
RN [4]

RP SEQUENCE FROM N.A.  
RA Onaka H., Taniguchi S., Igarashi Y., Furumai T.;  
RT "Cloning of the biosynthetic gene cluster of rebeccamycin, an  
RT anticancer agent, from *Lechevalieria aerocolonigenes*  
RT ATCC39243.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AF534707; AAC01211.1; -;  
DR EMBL; AB090952; BAC10677.1; -;  
DR EMBL; AJ141459; CAC93717.1; -;  
DR EMBL; AB071405; BAC15753.1; -;  
DR GO; GO:0004497; F:monooxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; p450; 1.  
DR PRINTS; PR00385; p450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR Heme; Monooxygenase; Oxidoreductase.

Query Match 28.4%; Score 582.5; DB 2; Length 397;  
Best Local Similarity 37.2%; Pred. No. 3.4e-34;  
Matches 139; Conservative 60; Mismatches 134; Indels 41; Gaps 9;  
QY 22 DAYHPPFWZESMRKADPVSDENQVWSVLYDDVKVVGDKEL-----PSSCMPQQT 74  
DB 13 ELADPVEVREYLTGDFVHNGE--AMYVFSYDGVAVHLTSRDYGRGPGGRATPIPPSH 70

QY 75 SSGT-----NSLIMNDPPKTKIRSVVNKAFTRPVKMQWEPRIQETIDELTQKQGRSEFD 130  
DB 71 DILSRIVENWLVLDPPRHTLSLLAKESPAVVTGLRERVKXIGALLAGLDAGEID 130

QY 131 LVHDFSYPLPVIVISLGLVPSAHMEOFKAW-----SDLLVSTPKDSEAEKA----- 179  
DB 131 LVDFDFAAPILVISELLGVPA-----RLRSWFRRCVLDL-----QEASTARATNP 177

QY 180 -FLEERDKCEEELAAAFAGIIEKKNKP-QDIIISILVEAEETGKLSGELIPFCITLL 237  
DB 178 GALARADGAASLVEVFGG--ELGTRKPDDEIVALLVNAQRGEALTDEIVSTCVHLL 235

QY 238 VAGNETTTNLISNAMYISILETPGVYELRSHPELMPQAVEALFRAPAPVLRRIAKRDT 297  
DB 236 TAGHEITTNLISKVALLANPAAABPLAGLDVTPQVVEELNRFDPVQVYVTEWAHQDT 295

QY 298 EUGHLIKEGDMVLAFFVANSRDEAKFDRPHMFDIRRHNPPIAAGHIFCLGAPLARL 357  
DB 296 ALGGKPIRRDKVVLVLSANROPAAFAEPDRDLDRDRSRHCGFGLGIHYCLGAALART 355

QY 358 EANTALTSLSIAFP 371  
DB 356 EAEIGLSVLFTNFP 369

RESULT 14  
Q59523 PRELIMINARY; PRT; 397 AA.  
AC Q59523;  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE Cytochrome P450 107E1 (Mycinaicin biosynthesis protein mycG).

GN CYP107E1 OR MYCG.  
OS Micromonospora griseorubida.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Micromonosporineae; Micromonosporaceae; Micromonospora.  
OX NCBI\_TaxID=28040;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=All1725;  
RX MEDLINE=95107242; PubMed=7808395;  
RA Inouye M., Takada Y., Muto N., Horinouchi S., Beppu T.;  
RT "Characterization and expression of a p-450-like mycinamicin  
RT biosynthesis gene using a novel Micromonospora-Escherichia coli  
RT shuttle cosmid vector.";  
RL Mol. Genet. 245:456-464 (1994).

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; D16098; BAA03672.1; -;  
DR PIR; S51594; S51594.  
DR HSPF; Q00441; 10XA.

DR GO; GO:0004497; F:monooxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; p450; 2.  
DR PRINTS; PR00385; p450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR Heme; Monooxygenase; Oxidoreductase.

Query Match 28.3%; Score 581; DB 2; Length 397;  
Best Local Similarity 37.3%; Pred. No. 4.4e-34;  
Matches 134; Conservative 64; Mismatches 131; Indels 30; Gaps 8;  
QY 30 YESMRKADPVSDENQVWSVLYDDVKVVGDKELFSSCMPQQT-----SS 76  
DB 23 YGSLQETEPVSRVPPYGE--AWLVTRYEDVRAVLGDGRFVRG--PSMTRDEPTREPM 78

QY 77 IGNSIMNDPPKTKIRSVVNKAFTRPVKMQWEPRIQETIDELTQKQGRSE-FDLVHDF 135  
DB 79 VKGLLSMDPEHSRLRLVVKAFTRAEASLPRASEIAHELVDQMAATQPADLVAMP 138

QY 136 SYPLPVIVISELGLVPSAHMEOFKANSCLIVSTPKDSEAEKAFLEERDKCEEELAAFF 195  
DB 139 ARQLPVVVICELGLVPSADHDFRMSGAFSLTAEVTAEMQEA-----AEQAYAYM 190

QY 196 AGIIEKKNKPEQDIISILVEAEETGKLSGELIPFCTLLVAGNETTNLISNAMYSI 255  
DB 191 GDLIDRRKKEPTDGLVSALVQARDQDLSQEQLLDLAIGLLVAGYESTTQIADFWILL 250

QY 256 LETPGVYELRSHPELMPQAVEALFRAPAPV---LRIAKRDETEIGGHLIKEGDMVLA 312  
DB 251 MTRPELRQLLDRELIPSAVEELTRW-VPLGVGTAFPRYAVEDVTLRGVTRAGEPVLA 309

QY 313 FVANSRDEAKFDRPHMFDIRRHNPPIAAGHIFCLGAPLARLEANIATLSLSIAFP 371  
DB 310 STGAANRDOAQFPDADRIDVDRTFNQHLGFGHVGHCIGAPLARVELQVALEVLQRLP 368

RESULT 15  
Q9CBE7 PRELIMINARY; PRT; 434 AA.  
AC Q9CBE7;  
DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE Putative cytochrome p450.  
GN MZ2088.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TW;  
RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream K.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 403:1007-1011(2001).  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; AL583924; CAC31043.1; -.  
DR PIR; C87170; C87170.  
DR HSSP; Q00441; LJIP.  
DR Leproma; ML2088; -.  
DR GO; GO:0004497; F:monooxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; P450; 1.  
DR PRINITS; PR00385; P450.  
KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.  
SQ SEQUENCE 434 AA; 47154 MW; 60A83B14B2769067 CRC64;

Query Match 28.3%; Score 581; DB 16; Length 434;  
Best Local Similarity 35.1%; Pred. No. 5e-34;  
Matches 136; Conservative 80; Mismatches 140; Indels 32; Gaps 6;

QY 8 QALQFALLNGKKNQDAYHPFMYESMRKDAVSF-DEENQVWSVFLYDD--VKVVVGDKK 64  
DB 35 QVLLRLLDPGTRD---PFPYRALIDYGMQLPGMPLTVFSSDCDDELRHPLSASD 91  
QY 65 LFSQCMPOQTSSIG-----NSIINDPPKHTKIRSVVNKAFTPRVMKQWEPRIQET 116  
DB 92 RLKATLAQAATAAGAEPRPFYASSPFMFLDPDPTRLRLKLVSKAFAPKVQALEGDIALLV 151  
QY 117 DELIQFGQRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQFKAWSDDLIVST----- 168  
DB 152 DSLIDKGAAGQFDVIADLAPLAVAVICRLGLGVPEADAPFGRVSRVALIVQSVDPFIT 211  
QY 169 --PKDKSEAEKAFLEERDKCEEELAAFPAGLIEBKXKPEODIISILVEAETGKLSG 226  
DB 212 GEPTTEERLRAGVRLRDYEQ-----LVKCRGTGEGEDLISRLIELDESGOQLTE 263  
QY 227 EELIPFCPLLIVAGMETTNLISNAMYSILETEGVYSRLSHPLMPQAVEALRFRAPA 286  
DB 264 EEIATCGLLLVAGHTTNVLIANAVLMLRNPQWKALSSNPQRAPLVVEETLRYDPAI 323  
QY 287 PVLRIAKRDEIGGHLIKEGDMVLAFVASANRDAKDRPHMFDIRRHNPPIAFGHCI 346  
DB 324 HLIQKVAADMTIGTTLTEGTMVLLAAANRDPAVYSRPDEFPDPRPSRHLAFVAGS 383  
QY 347 HFCIGAPLARLEANIATLSLISAPFME 374  
DB 384 HFCIGALARLEATVTLAISARFPQVQ 411

Search completed: May 28, 2004, 23:22:10

Job time : 225 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:22:17 ; Search time 352 Seconds  
(without alignments)  
314.156 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052

Sequence: 1 MNVLRNRQALQRLANGK.....SITPIENSVIYGLKSRVKM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2003000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670	32.7	410	14	US-10-205-032-6
2	649.5	31.7	404	14	US-10-214-446-40
3	642	31.3	416	9	US-09-861-289-39
4	642	31.3	416	9	US-09-860-846-39
5	642	31.3	416	10	US-09-988-384B-39
6	642	31.3	416	10	US-09-836-821-39
7	642	31.3	416	10	US-09-793-708-18
8	642	31.3	416	14	US-10-201-365-13
9	642	31.3	416	14	US-10-160-539-18
10	642	31.3	416	14	US-10-271-889-39
11	636	31.0	425	14	US-10-214-446-20
12	634.5	30.9	393	14	US-10-156-761-9525
13	623.5	30.4	392	14	US-10-214-446-32
14	622	30.3	457	14	US-10-156-761-11073
15	620.5	30.2	403	16	US-10-229-148B-20

16	599.5	29.2	388	14	US-10-156-761-13776
17	587.5	28.6	408	14	US-10-214-446-4
18	586	28.6	404	14	US-10-214-446-50
19	576.5	28.1	430	9	US-09-738-626-4117
20	571	27.8	418	12	US-10-389-647-559
21	570.5	27.8	405	14	US-10-156-761-12073
22	565	27.5	399	14	US-10-314-657-37
23	557	27.1	404	14	US-10-321-188-76
24	550	26.8	399	14	US-10-156-761-8126
25	550	26.8	432	14	US-10-145-415-10
26	549	26.8	429	14	US-10-145-415-14
27	549	26.8	430	14	US-10-145-415-30
28	548	26.8	430	14	US-10-145-415-95
29	548	26.7	418	16	US-10-461-194-118
30	547	26.7	430	14	US-10-145-415-8
31	547	26.7	430	14	US-10-145-415-20
32	544	26.5	430	14	US-10-145-415-12
33	540	26.3	399	14	US-10-156-761-3914
34	540	26.3	430	14	US-10-145-415-26
35	539	26.3	415	14	US-10-214-446-46
36	537.5	26.2	430	14	US-10-145-415-4
37	537.5	26.2	430	14	US-10-145-415-28
38	536	26.1	430	14	US-10-145-415-18
39	532	25.9	404	14	US-10-156-761-10431
40	531	25.9	430	14	US-10-145-415-24
41	531	25.9	430	14	US-10-145-415-32
42	530	25.8	426	14	US-10-145-415-34
43	530	25.8	430	14	US-10-145-415-2
44	530	25.8	475	14	US-10-145-415-22
45	528	25.7	415	16	US-10-461-134-88

#### ALIGNMENTS

RESULT 1  
US-10-205-032-6  
; Sequence 6, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staiffa, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-205  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-6

Query Match 32.7%; Score 670; DB 14; Length 410;  
Best Local Similarity 40.1%; Pred. No. 3.2e-51;  
Matches 143; Conservative 65; Mismatches 141; Indels 8; Gaps 2;  
QY 22 DAYHFPWVESRKDAPVDFDENQVSWFLDDVKKVGVGDKELFSS-----CMFQQTSS 75  
DB 19 DAQGLDMFAYNRKMPVSWDETRQAWHVFSRDYQTVTTNPLIFSSDFTSVFFVSELA 78  
QY 76 SI--GNSIINMPPKHTKIRSVMKAFTRVMKQWEPRIQETDLEIKQFQGRSEHFLVH 133  
DB 79 LIMGFTGGIDPPHAPLRKLVSAQFPRRAQWELRGQITADVLDQVROQDRDIAS 138  
QY 134 DFSYPLPVIVISGLVPSAHEQFKAUSDLLVSTPKKSEFAEKAFLBERKCEBELAA 193  
DB 139 DLAYPLFVTVIAELLGIPTKDHEKFEWDIILSNEGLEYVNPDPDFTETVGPABEWS 198  
QY 194 FRAGTIEEKRNKPEODIISILVEAETGKLSGELIPFCITILLVAGNTTNTLNINAM 253

```
Db 199 FCYAQIAKRAEPKODLISGLCAAAYDGRKLTDEEVNIVALLTAGHSSATLNLFL 258
Qy 254 SILETPGVYEELRSHPELMPQAEALRFRAPAVLRRIAKRDTEIGGHLIKEGDMVLAF 313
Db 259 VLEHFPQAAVRAERSLVGVIEELRVRSPNCIFRILNEDTDILGHPMRKGQWVIAW 318
Qy 314 VASANDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLEANIATLSISAF 370
Db 319 IASANRDETEVFTDPTDFDIRRESNKLAFHGHIHFCGLGAPLARLEAKVFNQTLDOF 375

RESULT 2
US-10-214-446-40
; Sequence 40, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-40

Query Match 31.7%; Score 649.5; DB 14; Length 404;
Best Local Similarity 38.3%; Pred. No. 2.1e-49;
Matches 143; Conservative 69; Mismatches 122; Indels 39; Gaps 8;

Qy 26 PEPWESMRKDAVPSPDENQ-----VMSVFLYDDVKVVGDKELRSC-- 69
Db 18 PYPFNEADGISLADAYEAREQGLLRVMAYCEPAWLATRYADARLVLDGRR-FSRAEG 76
Qy 70 ----MPQOTSIGNS-IINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIQKF 124
Db 77 ARHDEPRQSEGRDRSGILSMDDPDHTRTLTVAKAFTMHQVEKLRPAVRELADLIDKMV 136
Qy 125 GR-SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLIVSTPKKSEAEAKAFLEE 183
Db 137 ATGAPVDLVEEFALFPVGIVICQLLGVPEVDRPRFRAWSDAALST-----SSLTABE 188
Qy 184 RDKCEELAAFPAGIIEEKRNKPEQDIISILVEABETGEKLSGEEELIPFCTLLLVAGNET 243
Db 189 FRANGEEELRAYWGLIEDHRAPREDLITGLTEARPDRLTEQELVLDVLCVGLVAGHET 248
Qy 244 TTNLSNAMYSLTEPGVVEELRSHPELMPQAEALRFRAPAVP-----LRRIAKRDTE 298
Db 249 TATQIPNFVVTLDRPEQWNLREDPELVPTAVEELMRF--VFLSGSGASFPYATVEDVE 305
Qy 299 ICGHLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLE 358
Db 306 VGGTLVRAGEPVLVAVGAANRDPAREDAPEQLDLAREGNQHLGFGHGVHFCGLGAPLARLE 365
Qy 359 AMIALTSLISAPP 371
Db 366 LQELCALURRLP 378
```

```
RESULT 3
US-09-861-289-39
; Sequence 39, Application US/09861289
```

```
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-39

Query Match 31.3%; Score 642; DB 9; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

Qy 21 QD-AYHPPFWYEMRKDAP---VSEDEENQVNSVFLYDDVKVVGDKELRSC-- 71
Db 22 QDFAADPYPTYARAEAGFAHRVTRTEGDEWLVWGVYDRARAVLADPRFSKDMRNSTPL 81
Qy 72 -QOTSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIQKF---QGR 126
Db 82 TEAAALNNHMLSEDDPRHTRIRKLVARFTMRVRELLRFRVQEIVDGLVDMALAPDGR 141
Qy 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLIVSTPKKSEAEAKAFLEE 186
Db 142 A--DMESLAWP-LITVISELLGVPEPDRAAPRWMDAFVF--PDDPAQAQTAM----- 191
Qy 187 CEEELAAFPAGIIEEKRNKPEQDIISILVE-ABETGEKLSGEEELIPFCTLLLVAGNETTT 245
Db 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDGSRITSELLGMAHILLVAGHETTV 249
Qy 246 NLISNAMYSLTEPGVVEELRSHPELMPQAEALRFRAPAVLRRIAKRDTEIGGHLI 304
Db 250 NLIANGMYALLSHPDQIALALRADMTLLDGAVEELRYEGEVESATYRFFVPEVDLGTVI 309
Qy 305 KEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLEANIAT 364
Db 310 PAGDTVLVVLADAHRTPEPRFPDPRDIRDRDTAGHLAFHGHIHFCIGAPLARLEARI 369
Qy 365 SLISAFPHMECVSITPIE-----NSVIYGLKSEFVK 395
Db 370 ALLERCPDL-ALCVSPGELVWYVNPMPMIRGLKALPIR 404

RESULT 4
US-09-860-846-39
; Sequence 39, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
```



Db 370 ALLERCPDL-ALDVSFGLVWYVNPVIRGLKALPIR 404

## RESULT 7

US-09-793-708-18  
; Sequence 18, Application US/09793708  
; Publication No. US20030104597A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002121  
; CURRENT APPLICATION NUMBER: US/09/793,708  
; CURRENT FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: US 09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: US 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: US 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: US 09/073,538  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/846,247  
; PRIOR FILING DATE: 1997-04-30  
; PRIOR APPLICATION NUMBER: US 60/134,990  
; PRIOR FILING DATE: 1999-05-20  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-793-708-18

Query Match 31.3%; Score 642; DB 10; Length 416;  
Best Local Similarity 37.4%; Pred. No. le-48;  
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;  
  
QY 21 QD-AVHPFPWYEMRKDAP---VSFDEMCWWSVLYDDVKVVG---KELFSSCMP- 71  
DB 22 QDFAADPYTYARLRAEGPAHRVTPGDEWLVVGYDRARAVLADPRFSKDWNSTPL 81  
QY 72 -QOTSSIGNSIINMDPKTKIRSVNKAFTPRVMKOWEPRIQETDELIOKF----QGR 126  
DB 82 TEAEALNHMLSDPPRHLKLVAREFTMRVLRVLRVQEIYDGLVDMALAPDGR 141  
QY 127 SEFDLVHDFSYPLPVIVISELLGVPSAHMQPKAWSDLLVSTPKDSEAEKAFLEERDK 186  
DB 142 A--DLMESLAWPLPITVISELLGVPEPDRAAFRVWTDVAFV--PDPPAQQTAM----- 191  
QY 187 CBEELAAFFAGIIEKRNKPEQDIISILVE-ABETGEKLSGELIPFCTLLVAGNETTT 245  
DB 192 --AEMSGYLSRLIDSKRGQDGLLSALVTSDEDSRLTSELGMAHILLVAGHETTIV 249  
QY 246 NLISNAMYSILETGPVYEEIRSHPELMPQAVEALPRAPA-PVLRRIAKRDTTEIGHLI 304  
DB 250 NLIANGVYALLSHPDQLAALRADMTLLDGAVEEMLRVYEGVESATYRFPVPEVDLGTVI 309  
QY 305 XEGDMVLAFVASANRDEAKFDRPHMFDIRHPNPHIAFGHIFCIGAPLARLEANT 364  
DB 310 PAGDTVLVVLADAHRTPEFDRFDIRDRDTAGHLAFGHIFCIGAPLARLEARI 369  
QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSPVK 395  
DB 370 ALLERCPDL-ALDVSFGLVWYVNPVIRGLKALPIR 404

## RESULT 8

US-10-201-365-13  
; Sequence 13, Application US/10201365

Publication No. US20030148459A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR  
; FILE REFERENCE: 300622032103  
; CURRENT APPLICATION NUMBER: US/10/201,365  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: US 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: US 09/073,538  
; PRIOR FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-10-201-365-13

Query Match 31.3%; Score 642; DB 14; Length 416;  
Best Local Similarity 37.4%; Pred. No. le-48;  
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;  
  
QY 21 QD-AVHPFPWYEMRKDAP---VSFDEMCWWSVLYDDVKVVG---KELFSSCMP- 71  
DB 22 QDFAADPYTYARLRAEGPAHRVTPGDEWLVVGYDRARAVLADPRFSKDWNSTPL 81  
QY 72 -QOTSSIGNSIINMDPKTKIRSVNKAFTPRVMKOWEPRIQETDELIOKF----QGR 126  
DB 82 TEAEALNHMLSDPPRHLKLVAREFTMRVLRVLRVQEIYDGLVDMALAPDGR 141  
QY 127 SEFDLVHDFSYPLPVIVISELLGVPSAHMQPKAWSDLLVSTPKDSEAEKAFLEERDK 186  
DB 142 A--DLMESLAWPLPITVISELLGVPEPDRAAFRVWTDVAFV--PDPPAQQTAM----- 191  
QY 187 CBEELAAFFAGIIEKRNKPEQDIISILVE-ABETGEKLSGELIPFCTLLVAGNETTT 245  
DB 192 --AEMSGYLSRLIDSKRGQDGLLSALVTSDEDSRLTSELGMAHILLVAGHETTIV 249  
QY 246 NLISNAMYSILETGPVYEEIRSHPELMPQAVEALPRAPA-PVLRRIAKRDTTEIGHLI 304  
DB 250 NLIANGVYALLSHPDQLAALRADMTLLDGAVEEMLRVYEGVESATYRFPVPEVDLGTVI 309  
QY 305 XEGDMVLAFVASANRDEAKFDRPHMFDIRHPNPHIAFGHIFCIGAPLARLEANT 364  
DB 310 PAGDTVLVVLADAHRTPEFDRFDIRDRDTAGHLAFGHIFCIGAPLARLEARI 369  
QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSPVK 395  
DB 370 ALLERCPDL-ALDVSFGLVWYVNPVIRGLKALPIR 404

## RESULT 9

US-10-160-539-18  
; Sequence 18, Application US/10160539  
; Publication No. US20030162262A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/10/160,539  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07

```
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-160-539-18

Query_Match      31.3%; Score 642; DB 14; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYEMRKDAP---VSFDEENQVSVFLYDDVKVVGDP---KELFSSCMP- 71
Db 22 QDFAADPYTYARLRAEGPAHRVTRPGEDEVVLVVGVDARAVLADPRFSKDRNSTTPL 81
QY 72 -QOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVKQMPRIQEIITDELIQKF---QGR 126
Db 82 TEAAALNHNMLSDPPRHTLRKLVAEFTMKRVVLLPVRVQEIIVGLVDAMLAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSPKSKSEAEKAFLEERDK 186
Db 142 A--DLMESLAWPLPITVISELLGVPEPDRAAFRVWIDAFVF--PDDPAQAQTAM----- 191
QY 187 CEELAAFPAGIIEBKRNKPEQDIISILVE-AETGEKLSGSELIPFCTLLVAGNETTT 245
Db 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDEGSRLTSEELGMHILLVAGHETT 249
QY 246 NLISNAMYSLILETPGVYELRSHPELMPQAVEALRFRAPA-PVLRRIAKRDTETIGGHLI 304
Db 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEMLRYEGVESATYRFPVPEVDLGTVI 309
QY 305 KEGDWLAFVASANRDEAKFRPHMFDIRRHNPHTAFGHGHFCLGAPLARLEANIALT 364
Db 310 PAGDTVLVLADAHRTTPERPDPHFRDIRRDTAGHAFGHGHFCLGAPLARLEANIAVR 369
QY 365 SLISAPFPMCEVSTPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSFGELVWYFNPIMRGLKALPIR 404

RESULT 10
US-10-271-889-39
; Sequence 39, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; PRIOR FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-271-889-39
```

```
Query_Match      31.3%; Score 642; DB 14; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYEMRKDAP---VSFDEENQVSVFLYDDVKVVGDP---KELFSSCMP- 71
Db 22 QDFAADPYTYARLRAEGPAHRVTRPGEDEVVLVVGVDARAVLADPRFSKDRNSTTPL 81
QY 72 -QOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVKQMPRIQEIITDELIQKF---QGR 126
Db 82 TEAAALNHNMLSDPPRHTLRKLVAEFTMKRVVLLPVRVQEIIVGLVDAMLAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSPKSKSEAEKAFLEERDK 186
Db 142 A--DLMESLAWPLPITVISELLGVPEPDRAAFRVWIDAFVF--PDDPAQAQTAM----- 191
QY 187 CEELAAFPAGIIEBKRNKPEQDIISILVE-AETGEKLSGSELIPFCTLLVAGNETTT 245
Db 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDEGSRLTSEELGMHILLVAGHETT 249
QY 246 NLISNAMYSLILETPGVYELRSHPELMPQAVEALRFRAPA-PVLRRIAKRDTETIGGHLI 304
Db 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEMLRYEGVESATYRFPVPEVDLGTVI 309
QY 305 KEGDWLAFVASANRDEAKFRPHMFDIRRHNPHTAFGHGHFCLGAPLARLEANIALT 364
Db 310 PAGDTVLVLADAHRTTPERPDPHFRDIRRDTAGHAFGHGHFCLGAPLARLEANIAVR 369
QY 365 SLISAPFPMCEVSTPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSFGELVWYFNPIMRGLKALPIR 404

RESULT 11
US-10-214-446-20
; Sequence 20, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 6C/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-20

Query_Match      31.0%; Score 636; DB 14; Length 425;
Best Local Similarity 33.8%; Pred. No. 3.7e-48;
Matches 137; Conservative 79; Mismatches 145; Indels 44; Gaps 8;

QY 23 AVHPPFWYEMRKDAPV---SFDEENQVSVFLYDDVKVVGDKELF----- 56
Db 25 ASDPYAYANLREHAPVHRTLPSCVEAWLVTRYGDARQALADQRLSKNPAHHDSPHAK 84
QY 67 --SSCMPQOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVKQMPRIQEIITDELIQKF 124
Db 85 GKTGIPGERKAEALMTHLINIDPPDHTLRRLVSKAFTPRVVAEFTPRVQELTDLIDAFV 144
QY 125 GRSEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSPKSKSEAEKAFLE 181
```





```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11073
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11073

Query Match      30.3%; Score 622; DB 14; Length 457;
Best Local Similarity 33.3%; Pred. No. 7.4e-47;
Matches 137; Conservative 76; Mismatches 148; Indels 50; Gaps 9;

QY 23 AVHPFWMYEMRKDAPV----SPDEENQVWSVFLYDDVKKVGVGDKELFSSCMPQQTSSIGN 79
Db 55 ATDPYPAYAWLREHAPVHRTTLTPSGVEAWLVTRYADAKQALADARLSKNPVHSHSEDAPGK 114
QY 80 S-----IINMDPKHTKIRSVNKAFTPRVMKQWEPRIQITDELICKQF 124
Db 115 SGTGIPGERSANMLTHLLNDPDDTLRLRLVSKAFTPRVAEFAPRVQELTDHLIDQFA 174
QY 125 GRSEFDLVHDFSYPLVIVISLLGVPSAHMBOFKAWSDLLV---STPKDKSEAEKAF 181
Db 175 QTGSADLIHEFAPPLPIYALCDLGVPREQDDFDWAGMMIRHGGGPRGGVARSVK--- 231
QY 182 BERCKSEELAAFFAGLIEEKKNK-----PSQDIISILVEABETCKLSGEBLIPFCTL 235
Db 232 -----KMRGYLABLIHRKREALPADGPGEDLISGLIRASDHGEHLTENEAAAMCFV 283
QY 236 LLVAGNETTNNLISNAMYSILETPGVYEEL-----RSHPELMPQAVEEALRFPAPVLR 290
Db 284 LIFAGFETIINLIGNTYALLRNPPQARLQASIERGEQDLDTGIEELRYDGPVELAT 343
QY 291 -RIAKRDTIGGHLIKEGDMVLAFVASANRDEAKPDRPHMFDIRRHNPHEIAPFGHGHFC 349
Db 344 WRYATEPLDMGGQRIASGDPVLVWLAADRDPAFDEPDTLDSRDNQHLGYGHGHIYC 403
QY 350 LGAPLARLEANTALSLISAFPHMECVSITPIE-----NSVIYGLKSFYK 395
Db 404 IGAPLARLEGRATLTLRLRLPOLR--CAADPADLRWRGGLIMRGLNLEVE 453

RESULT 15
US-10-229-148B-20
; Sequence 20, Application US/10229148B
; Publication NO. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midcamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Streptomyces mycarofaciens
US-10-229-148B-20
```

```
Query Match      30.2%; Score 620.5; DB 16; Length 403;
Best Local Similarity 40.2%; Pred. No. 8.4e-47;
Matches 140; Conservative 55; Mismatches 144; Indels 9; Gaps 3;

QY 29 WYESMRKDAAPVSPDEENQVWSVFLYDDVKKVGVGDKELFSS-----CMPQQTSSIGNSI 81
Db 23 WFAFNTRTHHEVFWDESERHAWQVFRYDDLYLTVSNPQFFSSDFNEVMTTPPELEKVIQGT 82
QY 82 IN-MDPPKHTKIRSVNKAFTPRVMKQWEPRIQITDELICKQFQGRSEFDLVHDFSYPLP 140
Db 83 IGAIDPPAHGPMKLYSQAFTPRMAELEPRIRAVIQELLDVAVRGQETIDVVGDLVSLP 142
QY 141 VIVISELLGVPSAHMBOFKAWSDLLVSTPKDKSEAEKAFLEERDKCEELAEFFAGLIE 200
Db 143 VIVIAELLGIPSGDRDVFRCGWDVTLTNEGLEYPNLPDNFSETIAPALKEMTDYLLHQIH 202
QY 201 EKENKPEQDIISILVEAEETGEKLSGEBLIPFCTLILVAGNETTNNLISNAMYSILETPG 260
Db 203 AKREAPVDDLLISGLVQAEQGRKLTJVEIVNIIVALLTAGHVSSTLSNLFVLEENPQ 262
QY 261 VYBELRSHPELMPQAVEEALRFPAPVLRRIARRDTEIGGHLIKEGDMVLAFVASANRD 320
Db 263 ALADLRADRELVTGAVEETLRYSPFNNIIFRLKEDTDILGPEMKXQVVIAMQSANRD 322
QY 321 EAKFDRPHMFDIRRHNP-HIAFGHGHFCGAPLARLEANTALSLI 367
Db 323 PEHFPSPDTFDIRRSSSRHMAFGIGIHHLCLGAFARQEGKVVLNXL 370
```

Search completed: May 28, 2004, 23:54:59  
Job time : 353 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:12:17 ; Search time 92 Seconds  
(without alignments)  
222.216 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052  
Sequence: 1 MNVLRQALQALLGNKK.....SITPIENSVIYGLKSRVYM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/2/iaa/5A COMB.pap:\*
- 2: /cgm2\_6/ptodata/2/iaa/5B COMB.pap:\*
- 3: /cgm2\_6/ptodata/2/iaa/6A COMB.pap:\*
- 4: /cgm2\_6/ptodata/2/iaa/6B COMB.pap:\*
- 5: /cgm2\_6/ptodata/2/iaa/PTCUT COMB.pap:\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	642	31.3	416	3	US-09-320-878-18
2	642	31.3	416	3	US-09-105-537-39
3	642	31.3	416	4	US-09-141-908-13
4	642	31.3	416	4	US-09-657-440-18
5	606.5	29.6	399	4	US-08-765-907A-10
6	578	28.2	437	4	US-09-252-991A-17836
7	518.5	25.3	419	3	US-09-335-409-8
8	518.5	25.3	419	4	US-09-413-814-71
9	518.5	25.3	419	4	US-09-568-102-8
10	518.5	25.3	419	4	US-09-567-963-8
11	518.5	25.3	419	4	US-09-568-480-8
12	518.5	25.3	419	4	US-09-568-486-8
13	518.5	25.3	419	4	US-09-568-472-8
14	518.5	25.3	419	4	US-09-567-899-8
15	475	23.1	406	6	5212296-6
16	462	22.5	403	6	5212296-9
17	435.5	21.2	409	3	US-09-385-028-12
18	435.5	21.2	409	4	US-09-726-614-12
19	435.5	21.2	409	4	US-09-385-040-12
20	424	20.7	412	4	US-08-102-863-11
21	424	20.7	412	5	PCT-US92-10885-11
22	406.5	19.8	468	4	US-09-252-991A-32437
23	379	18.5	395	4	US-09-266-965-129
24	306	14.9	189	4	US-09-679-279-20
25	263	12.8	443	2	US-09-096-982-9
26	263	12.8	443	2	US-08-653-650A-9
27	262	12.8	422	2	US-09-096-982-5

28	262	12.8	422	2	US-08-653-650A-5	Sequence 5, Appli
29	262	12.8	474	2	US-09-096-982-8	Sequence 8, Appli
30	262	12.8	474	2	US-08-653-650A-8	Sequence 8, Appli
31	258	12.6	422	1	US-08-396-218-2	Sequence 2, Appli
32	258	12.6	422	1	US-08-760-116-2	Sequence 2, Appli
33	246	12.0	444	3	US-09-413-814-95	Sequence 95, Appli
34	246	12.0	444	3	US-09-413-814-96	Sequence 96, Appli
35	213.5	10.4	507	1	US-08-457-274A-23	Sequence 23, Appli
36	213.5	10.4	507	5	PCT-US95-05758-23	Sequence 23, Appli
37	209	10.2	503	3	US-09-144-367-2	Sequence 2, Appli
38	207.5	10.1	382	3	US-09-320-878-7	Sequence 7, Appli
39	207.5	10.1	382	4	US-09-141-908-7	Sequence 7, Appli
40	207.5	10.1	382	4	US-09-657-440-7	Sequence 7, Appli
41	207.5	10.1	402	3	US-09-105-537-22	Sequence 22, Appli
42	207.5	10.1	3782	3	US-09-105-537-4	Sequence 4, Appli
43	207	10.1	501	3	US-09-111-730-1	Sequence 1, Appli
44	200.5	9.8	467	4	US-09-126-420A-17	Sequence 17, Appli
45	199	9.7	501	3	US-08-906-791-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-320-878-18  
; Sequence 18, Application US/09320878A  
; Patent No. 6117659  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT MARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30622002120  
; CURRENT APPLICATION NUMBER: US/09/320,878A  
; CURRENT FILING DATE: 1999-05-27  
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30  
; EARLIER APPLICATION NUMBER: 60/119,139  
; EARLIER FILING DATE: 1999-02-08  
; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22  
; EARLIER APPLICATION NUMBER: 60/087,080  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-320-878-18

Query Match	31.3%;	Score 642;	DB 3;	Length 416;
Best Local Similarity	37.4%;	Pred. No. 2.4e-54;		
Matches 148;	Conservative 77;	Mismatches 137;	Indels 34;	Gaps 12;
QY	21	QD-AVHPFWYEMRKDA2--VSFDEENQVMSVFLYDDVKVVGDD---KELFSSCMP- 71		
Db	22	QDFAADPYTYARLRAGFAHRVTPGDEVLVVGVDARAVLADPRFSKWNSTPL 81		
QY	72	-QOTSGNSIINMPKHTKIRSVVKNKFTPRVWKQWEPRIQETDEITQKF---QGR 126		
Db	82	TAEAAINHNMLESPRHTRLKLVAREFTWRVLELLPRVQEIYDGLVDAMLA2DGR 141		
QY	127	SFEDLVHDSYPLPVIVISSELLGVPSAHMEQPKAMSDLLVSTPKDSSEAEKAFLEERDK 186		
Db	142	A--DLMESLAWFLPTVISELLGVPEPDRAAFRVWTDFAVF--PDDPAQAQTAM----- 191		

```

QY 187 CEEELAAFFAGIIEEKNKPEQDIISILVE-ABETGKLSGEBLIPFCTILLVAGNETTT 245
DB 192 --AEMSGYLSRLIDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLISNAMYSLTETPGVYEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHI 304
DB 250 NLIANGMYALLSHPDQALALRAJMTLLDGAVEEMLYEGVESATYRFFVPEVDLDGTVI 309
QY 305 KEGDMVLAFAVANROEAKFDRPHMEDIRRHNPHTAFGHGIFCLGAPLARLEANIALT 364
DB 310 PAGDTVLVVLADAHRTPERFPDPRHEDIRDTAGHLAFGHGIFCLGAPLARLEARI 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
DB 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

```

## RESULT 2

```

US-09-105-537-39
; Sequence 39, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Ashley, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-39

```

```

Query Match 31.3%; Score 642; DB 3; Length 416;
Best Local Similarity 37.4%; Pred. No. 2.4e-54;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

```

```

QY 21 QD-AVHFFPWYESMRKDP---VSPDEENQVNSVFLYDDVKVVGD----KELFSSCMP- 71
DB 22 QDFAADPYTYARLRAEGRAHVRTEGDEVWLVVGVDRARAVLADPRFSKDRNSTTPL 81
QY 72 -QOTSSIGNSIINMDPPKTKIRSVVNKAFTPRVMKOWEPRIOEITDELIOKF-----QGR 126
DB 82 TEAAALNHNMLESDPPRHTLRKLVARFTMRVVELLRPRVQEIVDGLVDMALAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQKAWSDLLVSTPKDSEAEKAPLEERDK 186
DB 142 A--DLMESLAWPLPIVISELLGVPEPDRAAFRVWTDAFV--PDPPAQOATM----- 191
QY 187 CEEELAAFFAGIIEEKNKPEQDIISILVE-ABETGKLSGEBLIPFCTILLVAGNETTT 245
DB 192 --AEMSGYLSRLIDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLISNAMYSLTETPGVYEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHI 304
DB 250 NLIANGMYALLSHPDQALALRAJMTLLDGAVEEMLYEGVESATYRFFVPEVDLDGTVI 309
QY 305 KEGDMVLAFAVANROEAKFDRPHMEDIRRHNPHTAFGHGIFCLGAPLARLEANIALT 364
DB 310 PAGDTVLVVLADAHRTPERFPDPRHEDIRDTAGHLAFGHGIFCLGAPLARLEARI 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
DB 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

```

## RESULT 3

```

US-09-141-908-13

```

```

; Sequence 13, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: Ashley, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 30622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-13

```

```

Query Match 31.3%; Score 642; DB 4; Length 416;
Best Local Similarity 37.4%; Pred. No. 2.4e-54;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;
QY 21 QD-AVHFFPWYESMRKDP---VSPDEENQVNSVFLYDDVKVVGD----KELFSSCMP- 71
DB 22 QDFAADPYTYARLRAEGRAHVRTEGDEVWLVVGVDRARAVLADPRFSKDRNSTTPL 81
QY 72 -QOTSSIGNSIINMDPPKTKIRSVVNKAFTPRVMKOWEPRIOEITDELIOKF-----QGR 126
DB 82 TEAAALNHNMLESDPPRHTLRKLVARFTMRVVELLRPRVQEIVDGLVDMALAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQKAWSDLLVSTPKDSEAEKAPLEERDK 186
DB 142 A--DLMESLAWPLPIVISELLGVPEPDRAAFRVWTDAFV--PDPPAQOATM----- 191
QY 187 CEEELAAFFAGIIEEKNKPEQDIISILVE-ABETGKLSGEBLIPFCTILLVAGNETTT 245
DB 192 --AEMSGYLSRLIDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLISNAMYSLTETPGVYEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHI 304
DB 250 NLIANGMYALLSHPDQALALRAJMTLLDGAVEEMLYEGVESATYRFFVPEVDLDGTVI 309
QY 305 KEGDMVLAFAVANROEAKFDRPHMEDIRRHNPHTAFGHGIFCLGAPLARLEANIALT 364
DB 310 PAGDTVLVVLADAHRTPERFPDPRHEDIRDTAGHLAFGHGIFCLGAPLARLEARI 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
DB 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

```

## RESULT 4

```

US-09-657-440-18
; Sequence 18, Application US/09657440
; Patent No. 6503455
; GENERAL INFORMATION:
; APPLICANT: Ashley, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

```

```

; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-10

Query Match      29.6%; Score 606.5; DB 4; Length 399;
Best Local Similarity 38.0%; Pred. NC. 6.9e-51;
Matches 136; Conservative 66; Mismatches 133; Indels 23; Gaps 7;

QY 29 WYSEMEKDAVPSEDEENQWVSFLDYDVKKVGDKELFSSCMPE---QTSSIGNSIIN-M 84
DB 19 WLREMKHHHPVHEDEVG-AHFWRHADVUTVASDPGVYSSQLSRLRPGSQALSEQLISVI 77

QY 85 DPKPKTKIRSVVNKAFTPRVQWQWEPRIQIETDELIOKQGRSEFDLVDFSYPLPVIVI 144
DB 78 DPMHRTLRRLYSQAFTPRVADLEPRVTLAGQLLDADVQ-DTDFLDVADFAYPLPVIVI 136

QY 145 SELLGPSAHMQFOKAWSDLLVSTPKDSEEAQKAFLEERDKCEE------LAA 193
DB 137 AELGVPFPADRTLFRSWSRML-----QMOVDADPMQFGDDADEYQRLVKPEPMRAMHA 191

QY 194 FPAIGIIEKRNKPEQDIISILVEAETGEKLGSEELIPFCTLLVAGNETTTNLIISAMY 253
DB 192 YLHDHVTDRRAPNDLISALVAARVEGERLTDEQIVFEGALLMAGHVSTSMLLGNTVL 251

QY 254 SILETGVTEELRSHPELMPQAVEALRPRAPVPRRIAKRDTTEIGGHLIKEGDMVLAF 313
DB 252 CLKDHPRAEAAARASRLIPALIEEVLRLRPPIITWARTVKDTVLAGTTIPAGRNVVPS 311

QY 314 VASANDEAKFDPMPFMDRRHPNPHIATFGHGHFCLGAPLARLEANIATLSISAPP 371
DB 312 LLSANNDEQVFTDPHLDLARE-GRQIATFGHGHYICLGAPLARLEORIALLEALFDRFP 368

RESULT 6
US-09-252-991A-17836
; Sequence 17836, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17836
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17836

Query Match      28.2%; Score 578; DB 4; Length 437;
Best Local Similarity 33.4%; Pred. No. 4.9e-48;
Matches 131; Conservative 76; Mismatches 147; Indels 38; Gaps 8;

QY 2 NVLNRQALQORALLGNKKODAYHPFPWYFESMRKDA---PVSEDEENQWVSFLYDDVKK 58
DB 19 NVPDRLKRLGEEILS-----PULAL--YDGLVOGAPRPAHRAEHPVMVVTVRYDARK 70

QY 59 VVGD-----KELFSSCMPQQTSSIGNSI-----INMDPKHTKIRSVNKAFTPRV 104
DB 71 VLNHGVRDRDARQAALYAKRTGSPRAGIGEGUSHMLNLDPPDHPRLKSLVGAFETPRQ 130

QY 105 MKQWEPRIQIETDELIOKQGRSEFDLVDFPSYPLPVIVISELLGVPSAHMQFOK-AWSD 163
DB 131 VERLQPHIERITTELLDANAGREQADLMADFAIPLATIVIFELGLGIPAEAREHARQSW-- 188

QY 164 LLVSTPKDSEEAQKAFLEERDKCEBELAAFFAGIIEBKRNKSEQDIISILVEAETGEK 223
DB 189 -----ERQAEILSPESAQALAAQVDYLVLL-ENAKRQPADVYSGLVQAADSEGG 239

```

```
QY 224 LSGBELPCTLLVAGNETTTLNISNAMYSLTETPGVVEELRSHPELMPQAVEALRFR 283
DB 240 LSEAEVSWAHLMMSGFTTMMIGNALVTLVNEQALRAQPELLPNAEELVRHD 299
QY 284 APAPV-LRRIAKRDTIGGHLKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAF 342
DB 300 SPVRASMLRFTVEDVGLQVTPAGBYILVSNLTANHDAERFDDPRDLTRNTDGLGV 359
QY 343 GHGIFCLGAPLARLEANTALSLISAPFME 374
DB 360 GFGVHCVCASLARLEGRATQRLARFPDLQ 391

RESULT 7
US-09-335-409-8
; Sequence 8, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-8

Query Match 25.3%; Score 518.5; DB 3; Length 419;
Best Local Similarity 32.6%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;

QY 23 AYHPPFWYESMRKADPVSEDEENQVMSVFLYDDVKKVWGDKEL-----FSSCM 70
DB 25 AEDPPFAIERLEAREATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSRREWESEAYSSAI 84
QY 71 PQOTSIGNSIINMDPPKTKIRSVVNKAFTPRVWKQWEPRIQEIITDELIOKQGRSEFD 130
DB 85 PELSDMKXYGLFGLPEDHARVKLVNPSFTSRAIDLRAEIQRTVDQLLDARSQGEED 144
QY 131 LVHDFSYPPLVIVISSELLGVPSAHMEQKAWSD-----LLVSTPKDKSEAEKAFLEERD 185
DB 145 VVRDYAEGIPMRAISALLKVPACDEKFRFGSATARALGVGLVPOVDEETKTLVA--- 200
QY 186 KCEELAAFFAGIIEKRNKPEODIISILVBAETGKLSGELIPFCTLLIVAGNETTT 245
DB 201 SVTEGALLHDVLDERRRNPLENDVLTMLQAEADGSRSLTKELVALVGAIIAAGTDTTI 260
QY 246 NLISNAMYSLTETPGVVEELRSHPELMPQAVEALRFRAPVLR-----RIAKRDTIEGG 301
DB 261 YLIAFAVNLRSPEALELVKAEPLMRNALDEVLRP---DNILRTGTVRFARQDLEYCG 317
QY 302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFCLGAPLARLEANI 361
DB 318 ASIKKGEMVFLIIPSAIRDGTVFSPDVFVRRDTCGASLAYGRGPHVCFGVSLARLEAEI 377
QY 362 ALTSLISAPFMECVSITPI 381
DB 378 AVGTIFRFRPEMK-LKETPV 396

RESULT 8
US-09-413-814-71
; Sequence 8, Application US/09413814
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
```

```
Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-71

Query Match 25.3%; Score 518.5; DB 3; Length 419;
Best Local Similarity 32.6%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;

QY 23 AYHPPFWYESMRKADPVSEDEENQVMSVFLYDDVKKVWGDKEL-----FSSCM 70
DB 25 AEDPPFAIERLEAREATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSRREWESEAYSSAI 84
QY 71 PQOTSIGNSIINMDPPKTKIRSVVNKAFTPRVWKQWEPRIQEIITDELIOKQGRSEFD 130
DB 85 PELSDMKXYGLFGLPEDHARVKLVNPSFTSRAIDLRAEIQRTVDQLLDARSQGEED 144
QY 131 LVHDFSYPPLVIVISSELLGVPSAHMEQKAWSD-----LLVSTPKDKSEAEKAFLEERD 185
DB 145 VVRDYAEGIPMRAISALLKVPACDEKFRFGSATARALGVGLVPOVDEETKTLVA--- 200
QY 186 KCEELAAFFAGIIEKRNKPEODIISILVBAETGKLSGELIPFCTLLIVAGNETTT 245
DB 201 SVTEGALLHDVLDERRRNPLENDVLTMLQAEADGSRSLTKELVALVGAIIAAGTDTTI 260
QY 246 NLISNAMYSLTETPGVVEELRSHPELMPQAVEALRFRAPVLR-----RIAKRDTIEGG 301
DB 261 YLIAFAVNLRSPEALELVKAEPLMRNALDEVLRP---DNILRTGTVRFARQDLEYCG 317
QY 302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFCLGAPLARLEANI 361
DB 318 ASIKKGEMVFLIIPSAIRDGTVFSPDVFVRRDTCGASLAYGRGPHVCFGVSLARLEAEI 377
QY 362 ALTSLISAPFMECVSITPI 381
DB 378 AVGTIFRFRPEMK-LKETPV 396

RESULT 9
US-09-568-102-8
; Sequence 8, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
```

Matches	124;	Conservative	72;	Mismatches	155;	Indels	29;	Gaps	6
Qy	23	AYHPFPWYESMRKADPVSFDEENQVMSVFLYDDVKVVGDKEL-----FSSCM	70						
Db	25	AEDPFPALERLEATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSREWESSAEYSSAI	84						
Qy	71	PQOTSSIGNSIINMDPPKHTKTSVWNKAFTRPVWKQWEPRIQETIDELIQFGSRSEFD	130						
Db	85	PELSDMKKYGFLGPPEDHARVKLVNPSFTSRAIDLRLAEIQRTVDQLLDARSQGEED	144						
Qy	131	LVDHDSYPLPVIVISSELLGVSAHMEQKAWSD-----LLVSTPKDKSSEAEKAFLEERD	185						
Db	145	VVRDYAEGIFPMRAISALLKVPACDEKFRFGSATARALGVGLVPQVDEETKLVA----	200						
Qy	186	KCEELAAFPAGIIEBKRNKPPQDIISILVEABETGEKLSGSELI PFCTLLVAGNETTT	245						
Db	201	SVTEGLALLHDVLDERRRPLENDVLTMLQAEADGSRSLSTKELVALGVAINAGTDTTI	260						
Qy	246	NLISNAMYSLTEPGVYVELRSHPELMPQAVEALRFPAPVLR-----RIAKROTEIGG	301						
Db	261	YLIAFAVINLRSPALELWKAEPGLMRNALDEVLRF--DNLRIGTVRFARQDLEYCG	317						
Qy	302	HLIKEGDWLAFVASANRDEAKEDRPHMDIRHPHPHIAFGHGIFCCGAPLARLEANI	361						
Db	318	ASIKKGMVFLIPSAIRDGTVESREDFVDVRDGTGASLAYGRGPHPCPQVSLARLEANI	377						
Qy	362	ALTSLSIAFPHMECVSITPI	381						
Db	378	AVGTIFRRPEMK-LKETPV	396						
RESULT 11									
US-09-568-480-8									
; Sequence 8. Application US/09568480									
; Patent No. 6355458									
; GENERAL INFORMATION:									
; APPLICANT: Schupp, Thomas									
; APPLICANT: Ligon, James									
; APPLICANT: Molnar, Istvan									
; APPLICANT: Zirkle, Ross									
; APPLICANT: Cyr, Devon									
; APPLICANT: Goerlach, Joern									
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES									
; FILE REFERENCE: 4-30582A									
; CURRENT APPLICATION NUMBER: US/09/568,480									
; CURRENT FILING DATE: 2000-05-10									
; PRIOR APPLICATION NUMBER: 09/335,409									
; PRIOR FILING DATE: 1999-06-17									
; NUMBER OF SEQ ID NOS: 30									
; SOFTWARE: PatentIn Ver. 2.0									
; SEQ ID NO 8									
; LENGTH: 419									
; TYPE: PRT									
; ORGANISM: Sorangium cellulosum									
US-09-568-480-8									
Query Match 25.3%; Score 518.5; DB 4; Length 419;									
Best Local Similarity 32.5%; Pred.No.3.1e-42;									
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;									
Qy	23	AYHPFPWYESMRKADPVSFDEENQVMSVFLYDDVKVVGDKEL-----FSSCM	70						
Db	25	AEDPFPALERLEATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSREWESSAEYSSAI	84						
Qy	71	PQOTSSIGNSIINMDPPKHTKTSVWNKAFTRPVWKQWEPRIQETIDELIQFGSRSEFD	130						
Db	85	PELSDMKKYGFLGPPEDHARVKLVNPSFTSRAIDLRLAEIQRTVDQLLDARSQGEED	144						
Qy	131	LVDHDSYPLPVIVISSELLGVSAHMEQKAWSD-----LLVSTPKDKSSEAEKAFLEERD	185						
Db	145	VVRDYAEGIFPMRAISALLKVPACDEKFRFGSATARALGVGLVPQVDEETKLVA----	200						
Qy	186	KCEELAAFPAGIIEBKRNKPPQDIISILVEABETGEKLSGSELI PFCTLLVAGNETTT	245						



```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-8

Query Match      25.3%; Score 518.5; DB 4; Length 419;
Best Local Similarity 32.6%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;

QY 23 AYHPPWYEMRKDAPVDFDERNOVNSVFLYDDVKVVGDKEL-----FSSCM 70
DB 25 AEDPFAIERLEATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSRREWESSAYSSAI 84
QY 71 PQTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVNMQWEPRIQETIDELIQFQGRSEPD 130
DB 85 PELSDMKYGLFGLPPEDHARVKLVNPSFTSRADLLRAEIQRTVDQLDARSQGERFD 144
QY 131 LVHDFSYPLVIVISELGVPSAHMEQKAWSD-----LLVSTPKDKSEAEKAFLEERD 185
DB 145 VVRDYAEGIPRAISALLKVPACEDKFRFGSATARALGVGLVPOVDDEETKLVA---- 200
QY 186 KCEELAAFFAGITTEKRNKPEQDIISILVEAETGKLSGELIPFCTLLLVAGNETT 245
DB 201 SVTEGLALLHVDLORRRNPLENDVLTMLLQEAAGSRLSTKELVALVCAIIAAGTDTTI 260
QY 246 NLISNAMYSILETGVTYELRSHPELMPQAVEEALRFRAPAVLR-----RIAKRDTTEIGG 301
DB 261 YLIAFAVLNLLRSPEALELVKAEPLMRNALDELVR-----DNILRIGTVRPARQDLEYCG 317
QY 302 HLIKEGDMVLAFAVSAWRDEAKFDRPHMFDIRHPNPHIAFGHGHFCLGAPLARLEAM 361
DB 318 ASIKKGENVFLIIPSAIRDGTVFSRPFVDFVRDRTGASLAYGRGPHVCPGVSLARLEAEI 377
QY 362 ALTSLISAFPHMECVSITPI 381
DB 378 AVGIFRRFPENK-LKETPV 396

RESULT 15
5212296-6
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE HARDER, PATRICIA A.;LETO, KENNETH
; J.; O'KEEFE, DANIEL P.;OMER, CHARLES A.;ROMESSER, JAMES A.
; TEPFERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:6;
; LENGTH: 406
5212296-6

Query Match      23.1%; Score 475; DB 6; Length 406;
Best Local Similarity 32.2%; Pred. No. 5.4e-36;
Matches 125; Conservative 68; Mismatches 159; Indels 36; Gaps 10;

QY 19 NKQDAYHPPWYEMRKDAP-----VSPDEENQVMSVELYDDVKVVGDKEL----- 65
DB 19 NRSCEYQLPDGYAQLR-DTFCPLHRVTLYDGRQAWVTYKHEARKLLIGDPRJSSNRDTDN 77
```

```
QY 66 FSSCMP--QOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVNMQWEPRIQETIT---DEL 119
DB 78 FPATSPRFEAVRESPOAFIGLDPPEHGTTRRTISFTVKRIKGMREFEVEVHGLDEM 137
QY 120 IQKFQGRSEFDLVHDFSYPLVIVISELGVPSAHMEQKAWSDLLVSTPKDKSEAEKA 179
DB 138 L---AAGPTADLVSOQFALPVPMSVICRLGLGVPIADHEFFQDASKRLV-----OSTDAQSA 189
QY 180 FLEERDKCEELAAFPAGITTEKRNKPEQDIISILVEAETGKLSGELIPFCTLLLVLA 239
DB 190 LTAEND-----LAGYLDGLITQFTEPGAGLVGALVADQLANGEIDREELISTAMLLIA 244
QY 240 GNETTNLISNAMYSILETGVTYELRSHPELMPQAVEEALRFRAPAVLR---RIAKRDT 298
DB 245 GHETTASMTSLSVITILLDHPQYAAALRADRSIIVPGAVEELLRYLAIDAGGRVATADIE 304
QY 299 IGHHLIKKGDVLAFAVSAWRDEAKFDRPHMFDIRHPNPHIAFGHGHFCLGAPLARLE 358
DB 305 VEGHLIRAGEGVIVVNSIANRDGTVYEDPDALDIHRSARHHLAFGRGVHQCCLGQNLARLE 364
QY 359 ANIALTSLISAFPHMECVSITPIENSIV 386
DB 365 LEVILNALMDRVPTLRLA--VPVEQLVL 390

Search completed: May 28, 2004, 23:25:34
Job time : 93 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 21:28:11 ; Search time 437 Seconds  
(without alignments)  
12397.578 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtttaaccgccc.....gcttcggtgaaatgttaa 1191

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubnpa/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/1/pubnpa/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/1/pubnpa/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/1/pubnpa/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/1/pubnpa/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/1/pubnpa/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/1/pubnpa/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/1/pubnpa/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/1/pubnpa/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/1/pubnpa/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/1/pubnpa/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/1/pubnpa/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/1/pubnpa/US09\_NEW\_PUB.seq\*
- 14: /cgn2\_6/ptodata/1/pubnpa/US10A\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/1/pubnpa/US10B\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/1/pubnpa/US10C\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/1/pubnpa/US10\_NEW\_PUB.seq\*
- 18: /cgn2\_6/ptodata/1/pubnpa/US60\_NEW\_PUB.seq\*
- 19: /cgn2\_6/ptodata/1/pubnpa/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	425	35.7	966	9	US-09-974-300-2190
2	242	20.3	1213	9	US-09-974-300-2182
3	149	12.5	1233	15	US-10-205-032-7
C	149	12.5	60196	15	US-10-205-032-1
5	135.4	11.4	1197	15	US-10-156-761-576
C	135.4	11.4	9025608	15	US-10-156-761-1
7	121	10.2	84428	17	US-10-229-1488-1
8	120.4	10.1	1215	15	US-10-214-446-39
9	117.8	9.9	1197	15	US-10-156-761-2364
C	117.8	9.9	9025608	15	US-10-156-761-1
11	115.4	9.7	1215	15	US-10-156-761-4523
12	114.2	9.6	1164	15	US-10-156-761-6226
13	114.2	9.6	1248	15	US-10-214-446-55
14	112.8	9.5	1290	9	US-09-738-626-617

C	15	112.8	9.5	3309400	9	US-09-738-626-1	Sequence 1, Appli
	16	111.4	9.4	1257	13	US-10-389-647-206	Sequence 236, App
	17	110.2	9.3	1155	9	US-09-974-300-2185	Sequence 2185, Ap
	18	106.2	8.9	1179	15	US-10-156-761-1975	Sequence 1975, Ap
	19	106	8.9	1215	15	US-10-214-446-49	Sequence 49, Appl
C	20	105.4	8.8	82393	15	US-10-080-170-645	Sequence 645, App
	21	103	8.6	215	9	US-09-974-300-2242	Sequence 2242, Ap
	22	97.6	8.2	1191	15	US-10-156-761-7165	Sequence 7165, Ap
	23	96.8	8.1	1371	15	US-10-156-761-3523	Sequence 3523, Ap
	24	95.4	8.0	1278	15	US-10-214-446-19	Sequence 19, Appl
	25	93	7.8	1239	15	US-10-214-446-35	Sequence 35, Appl
	26	92.4	7.8	1209	15	US-10-205-032-9	Sequence 9, Appli
	27	92.4	7.8	1248	15	US-10-156-761-408	Sequence 408, App
	28	92.4	7.8	60196	15	US-10-205-032-1	Sequence 1, Appli
	29	92.4	7.8	100000	15	US-10-156-761-15103	Sequence 15103, A
	30	92.2	7.7	1233	15	US-10-156-761-826	Sequence 826, App
	31	90.4	7.6	135638	15	US-10-314-657-1	Sequence 1, Appli
C	32	89.6	7.5	86941	17	US-10-461-194-2	Sequence 2, Appli
	33	88.8	7.5	1233	10	US-09-953-348-24	Sequence 24, Appl
	34	88.8	7.5	1233	15	US-10-267-255-24	Sequence 24, Appl
	35	88.8	7.5	12249	10	US-09-953-348-74	Sequence 74, Appl
	36	88.8	7.5	12249	15	US-10-267-255-74	Sequence 74, Appl
	37	88.8	7.5	18331	10	US-09-953-348-96	Sequence 96, Appl
	38	88.8	7.5	18331	15	US-10-267-255-96	Sequence 96, Appl
	39	87.6	7.4	1200	13	US-10-282-122A-25888	Sequence 25888, A
C	40	87.6	7.4	71989	13	US-09-939-964-1	Sequence 1, Appli
	41	86.8	7.3	536165	10	US-10-156-761-7447	Sequence 7447, Ap
	42	86.6	7.3	1248	15	US-10-156-761-2153	Sequence 2153, Ap
C	43	85.8	7.2	1263	15	US-10-156-761-2153	Sequence 1, Appli
	44	84.8	7.1	85692	17	US-10-461-194-1	Sequence 22, Appl
C	45	84.4	7.1	1224	10	US-09-953-348-22	

#### ALIGNMENTS

#### RESULT 1

US-09-974-300-2190  
; Sequence 2190, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Herka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/05/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2190  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-2190

Query Match 35.7%; Score 425; DB 9; Length 966;  
Best Local Similarity 65.4%; Pred. No. 9.7e-127;  
Matches 623; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy	230	TTGGAATTCATCATTAACATGGACCGCCGCAACCATCAAAAATTCGTCAGTCGCA	289
Db	7	TTGGAATTCATCATTAACATGGACCGCCGCAACCATCAAAAATTCGTCAGTCGCA	66
Qy	290	ACAAAGCCTTTACTCCGGCGGTGATGAAGCAATGGAAACCGAGAATTCAGAAATACACAG	349
Db	67	ACAAAGCCTTTACTCCGGCGGTGATGAAGCAATGGAAACCGAGAATTCAGAAATACACAG	126
Qy	350	ATGAAGTATTCATCAAAAATTCAGGGCGCGCAGTGAAGTTTGACCTTGTTCACGATTTTCAT	409

```
Db 127 ATCAATTGCTCGCTGATGTCGCGGAGGAGAGATCGACCTTGTACAGATTTTCAT 186
Qy 410 ACCGCTTCGGTTATGTGATATCTGAGCTGTGGAGTGCCTTCAGCGCAGATGMAAC 469
Db 187 ATCCGTTGCCCGTAATCGTCAATTCGGAATGTCTCGCGCTTCCTTTGGTGATAGCATC 246
Qy 470 AGTTTAAACATGCTGCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAAGCTG 529
Db 247 ATTTCAGAAATGCTCGACCTGTTGTCAGCTTCGCGAAGAGCGATCGCGGAGATG 306
Qy 530 AAAAGACCTTTTGGAAAGACAGATAAGTGTGAGGAAGAACTGGCGCGCTTTTTCGCG 589
Db 307 TGAACAGTGGAAAAACATCAGGAGCAACAGGCCAAGAGAGAGCTGACCGCATCTTTGAAA 366
Qy 590 GCATCATAGAAAGAACCGAAACAAACCGGAACAGGATATTTCTATTTAGTGAAG 649
Db 367 AGATGATTCAGAGAAAGACAAACCTCGGCAATGACCTGATTTTCGCTTTGATCAAG 426
Qy 650 CGGAAGAAACAGCGGAGAGAGCTGCTCGGTGAGAGAGCTGATTTCCGTTTTCACGCTGCTGC 709
Db 427 CCGAACAAGAGGGAGAGCTGCTCCCTGATGAATTTGGTTCCATTTTGCATCTGCTCT 486
Qy 710 TGGTGGCGGAATGAACACCTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAG 769
Db 487 TGAATGCGGGGAATGAACACGACGCAATTTAGTTTCAAAATGCGGTCTACAGCATTTCTCG 546
Qy 770 AAACGCGAGCGGTTTACGAGGAACTCGGAGCCATCTGAACCTGATGTCCTCAGCGAGTGG 829
Db 547 AAACACCGGCGGTGATGACGAGCTCGCAGGCACTCTGAACTGATTTCCGAGGAGTGG 606
Qy 830 AGGAAGCTTTGGCTTTTCAGAGCGCGCGCCCGCTTTTTCAGGCGCATTTGCCAAGCGGATA 889
Db 607 AGGAAGCGTTCGCTTTTCGCGGCGCAGCGCGATGATTTGGCGCTTCGTTTAAACAGGATA 666
Qy 890 CGAGATCGGGGGGACCTGATTAAGAGAGTGATATGTTTGGCGTTTGGCGATCGG 949
Db 667 CGAGATCAGAGGATTAAGCTTTGAAAGAGGAGAGGTTGTGATCGCTTTCTCTTCTCTG 726
Qy 950 CAAATCGTATGAAGCAAAAGTTTGAAGACCGCAGCATTTTGTATTCGCGCGCATCCCCA 1009
Db 727 CCAACCGTATGAACGAATTTGAGAGGCGCGCAGCAATTTGATATTCACCGCATCCGA 786
Qy 1010 ATCCGATATTGCTTTGGCCAGCGCATCCATTTTTCCTTTGGGCGCCCGCTTGCCTGTC 1069
Db 787 ACCCGCATATCGGATTTGGGCAAGCGCATCCACTTTTGTGGGCGCCCGCTCGCAAGGC 846
Qy 1070 TTGAAGCAATATCGGTTAACTGCTTTGATTTCTGCTTTCTCATATGAGTGGTCA 1129
Db 847 TGAAGCGCGGATCGCCATTTGAGCGCTCTTAAAGCAGTATGCTCGATGGAAGAGCTTG 906
Qy 1130 GATCACTCCGATTGAAACAGTGTGATATACGATTAAGAGCTTCCGTTGTG 1182
Db 907 CGGTGCTGCGATGCGGACAGCAGCATGTACGCTTGAACATTTCCGCTG 959
```

## RESULT 2

```
US-09-974-300-2182
; Sequence 2182: Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10095.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/690,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
```

```
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 2182
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2182

Query Match      20.3%; Score 242; DB 9; Length 1213;
Best Local Similarity 53.6%; Pred. No. 3.7e-67;
Matches 550; Conservative 0; Mismatches 470; Indels 6; Gaps 2;

Qy 75 TCCTTTCCATGCTATGATTCGATGAGAAAGATGCGCCCTGTTCTTCTTGAATGAAGAAA 134
Db 92 TCCTTTCCGATTTATAACCGGATCAGAAAGAGCTCTCTCTGCTGCTTACGATGATGAGCG 151
Qy 135 CCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAAAGTTTGTTCGGGATAAAGATT 194
Db 152 AAATGCTTTGATATCTTTTCTTATGAGAGAGCTTCAATTCGTGTTAAAAAACCCGAGCT 211
Qy 195 GTTTCCTCAGTTGCTCGCGCAGCAGACAGCTCTATTTCGAAATTCATCTTAAACATGGA 254
Db 212 CTTCTC----TTCAAAACGCGCGGTATATGAA--GGAAAAAGTATATTAACAAATGA 265
Qy 255 CCGCGCAGCAGCATACAAAATCGTTTCAGTCTGTAACAAAGCCCTTACTCGCGCGCTGAT 314
Db 266 CCGCGCGCAGACACACAAAAATGAGAGCCATCGTTTAAATAAGCTTTTACGCGGAAAGCGGT 325
Qy 315 GAAGCAATGGAAACCGAGAAATCAAGAAATCAGAGATGAACCTGATTCAAAAATTCAGGG 374
Db 326 GAAGAGCTTGAACCGCATATCCGAGAGTGAAGCTTTTATTAAACGAGCGAACA 385
Qy 375 GCGCAGTGAATTTGACCTGTTTCAAGATTTTCAACCCGCTTCCGCTTATTTGATGATC 434
Db 386 GAAAGAAATTTGATGCTGTGAGCAGCTTGGCTGCTCTCTCTCCCTCATTTATCATCGC 445
Qy 435 TGAGCTGCTGGAGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCATGCTGATCTCT 494
Db 446 TGAATTTTTCAGGCTTTCCGCTGAAACCCCTCATGTTTAAACATTTATTCAGACATCTCT 505
Qy 495 GGTTCAGTACACCAAGATTAAGATGAAGAGCTGAAAAAGCCCTTTTGGAGAACCGAGA 554
Db 506 TGTCCGAGTGGCGAGACCGCTCGCTGAGCGCGCGAGCGATGTACAAACGAGCTGA 565
Qy 555 TAAATGTGAGGAAGAACTGCGCGCTTTTTCGCGCATCATAGAGAAAGCGAAACAA 614
Db 566 AGAAGCAATCGGTTTTCGCGGATTTTAAACAACTTATTCAGCGCGCGCAAAAAAGA 625
Qy 615 ACCGGAACAGGATATTTCTATTTTAGTGAAGCGGAAGAACAGCGCAGAGAGCTGTC 674
Db 626 GCCAAAGAGCAGCTGATTTTCGCTTTTTCGCGCGGGAAGTTGACGCGCAATTCGCTGAC 685
Qy 675 CGGTGAAGAGCTGATTCGCTTTTCGACGCTGCTGCTGGTGGCGCGGAAATGAAACCACTAC 734
Db 686 AGAAGAAAGCTGCTTCAATTTTTCGATCATCTTTTGGTGGCAGGCAATCAGACGCAAC 745
Qy 735 AAACCTGATTTTCAATGCGAGTACAGCATATTGAAACCCGAGGCGTTTACGAGAACT 794
Db 746 CAATCTGATCGCAACAGAGCTCGCTCTCTCACAGAAAGATAAAATCACACAGGAGCGCT 805
Qy 795 GCGCAGCCATCTGAACTGCTGCTCAGGAGTGGAGAGCCCTTGGCTTTCAGAGCGCC 854
Db 806 AAGACAAGATCCGCTCCCTGCTGCTTTCGTTTGAAGAAATGCTGCTGTTTATTCGCGC 865
Qy 855 GCGCCCGGTTTTCAGGCGCATTTGCCAAGCGGAGATCGGAGATCGGGGGGCACTGATTA 914
Db 866 CTGTCAAGCGATCGCGCGACGCGGCGCAGAGAGCTTGATTCGAGGCGCTGAGGATTC 925
Qy 915 AGAAGTGTATGCTTTTGGCGTTTGTGGCATCGGCAATCGTGAAGCAAGTTTGA 974
Db 926 AAAGGGTCTTACAGTGTATCAGCTGGGTGCTTCAGCGAATCGTGAACCTTAAGTTGA 985
Qy 975 CAGACCGCAGATGTTGATATCGCGCGCATCCCATCCCATATTCGCTTTCGCGCACGG 1034
```

Db 986 CGATCCTGACAGCTTCAAGCTTGATGCGMAATCAAAACCTCATATGAGAGTTTCGGCTTCGG 1045  
Qy 1035 CATCCATTTTGGCTTGGGGCCCGCTTGCCTGCTTGAAGCAAAATATGCGCTTAAAGCTC 1094  
Db 1046 CATCCATTTTGGCTTGGGGCTTCCCTCGCCGGCTTGAAGCAAAAGTTCGCTCGATTA 1105  
Qy 1095 TTTGAT 1100  
Db 1106 CTTGCT 1111

RESULT 3  
US-10-205-032-7  
; Sequence 7, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-20S  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 1233  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-7

Query Match 12.5%; Score 149; DB 15; Length 1233;  
Best Local Similarity 48.5%; Pred. No. 6.5e-37;  
Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;

Qy 239 CCATCAITAAATGACGCGCGGAGCATACAAAATCGTTTCAGTCGTGACAAAGCT 298  
Db 254 CCATCGCGGCGATGACCCGCGCGGCGCGCGCTGCGAAGCTGTGAGCCAGGCT 313  
Qy 299 TTACTCCGCGCTGATGAAGCAATCGGAACCGAGAAATCAAGAAATCAGATGAAGTGA 358  
Db 314 TCACCCCGCGGATCGCCAGATGAGTGCAGATCGGCGAGATCACCGCGAGCTGC 373  
Qy 359 TTCAAAAATTCAGGGCGCAGTGTGACCTTTCAGTTTTCATACCGCTTC 418  
Db 374 TCGACCAAGTACGCGACGAGCAGCGATCGACATCGCCAGCGACCTCGTACCGCTGC 433  
Qy 419 CGGTATGTGATATCTGAGTGTGCGAGTGCCTTCAGCGCAGATGAACAGTTTAAG 478  
Db 434 CGGTAGCGTATCGCGGCTGCTCGGCAATCCCAAGGATCAGAGAGTTCCCGC 493  
Qy 479 CATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGGCT 538  
Db 494 AGTGGTGGACATCATCTCAGCAAGAGGCTGAGTATCCCAACCTCCGAGAGCT 553  
Qy 539 TTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGCTTTTTCGCGCATCATAG 598  
Db 554 TCACCGAGACGCTGGGCGCCCGCTCGAGGAGTGTGTCGGAATTCGTATCGCCAGATCG 613  
Qy 599 AAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGAAGCGAAGAA 658  
Db 614 CCCAAGCGCGCGAACCAGGAAGAGCTGATCAGCGGCTCTGTGCGCGGAGGTCG 673  
Qy 659 CAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTTTGCACGCTGCTGTGTCGCG 718  
Db 674 ACGGCGCAAGCTGACCGACGAGGAGTGTCAACATGTCGCGCTGCTGTCTCACCGCG 733  
Qy 719 GAATGAAACCACTAGAACCTGATTTCAATGGATGTACAGCATATTAAGAACGCCAG 778  
Db 734 GGCACATCTCCAGCGCGCACGCTGTCTCAGCAACCTGTTCTGCTGTGAGGAGCACCGC 793  
Qy 779 GCGTTTACGAGAACTCGCGAGGCTATCTGAGTGTGATGATGCTCAGCACTGAGGAGCT 838

Db 794 AGSCACAGCGCGCGGTCCCGCGCGAGCCGCGAGCTCTCTGCGGGCGGTATCGAGGAGCGC 853  
Qy 839 TGGTTTTCAGAGCGCGCGGCCCGCTTTTCAGGCGCATTCGCAAGCGGATACGAGATCG 898  
Db 854 TGGCTACCGGTCCCGTTCACCTGATCTTCGGATCTGACGAGGACACCGACATCC 913  
Qy 899 GGGGCGACCTGATTAAAGAAAGTATATGTTTTCGGCTTGTGGCATCGGCAATCGTG 958  
Db 914 TCGGCGACCCCATGCGCAAGGCGCAGATGATGATCGCTCGATCGCTCGCGAACCAG 973  
Qy 959 ATGAAGCAAGTTTGACAGCGCGCATGTTGATATCGCGCCATCCCAATCCGATA 1018  
Db 974 ACACGAGGTGTTTCAGGACCCGACACCTTGACATCGACGCGAGTGAACAAGACC 1033  
Qy 1019 TTGCGTTTGGCCACGCGATCCATTTTTCCTTGGGCGCGCTTTCCTGCTTTGAAGCAA 1078  
Db 1034 TGGCGTTGCGCCACGCGATCCACACCTGCTGGCGCTTCTGCGCAGGCTGAGGCGA 1093  
Qy 1079 ATATC 1083  
Db 1094 AGGTC 1098

RESULT 4  
US-10-205-032-1/c  
; Sequence 1, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-20S  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 60196  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-1

Query Match 12.5%; Score 149; DB 15; Length 60196;  
Best Local Similarity 48.5%; Pred. No. 1e-35;  
Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;

Qy 239 CCATCAITAAATGACGCGCGGAGCATACAAAATCGTTTCAGTCGTGACAAAGCT 298  
Db 369 CCATCGCGGCGATCGACCCGCGCGCGCGCTGCGAAGCTGTGAGCGGCT 3549  
Qy 299 TTACTCCGCGCTGATGAAGCAATCGGAACCGAGAAATCAAGAAATCAGATGAAGTGA 358  
Db 3548 TCACCCCGCGGATCGCCAGATGAGTGCAGATCGGCGAGATCACCGCGAGCTGC 3489  
Qy 359 TTCAAAAATTCAGGGCGCAGTGTGACCTTTCAGTTTTCATACCGCTTC 418  
Db 3488 TCGACCAAGTACCGCACCGAGCAGATCGACATCGCAGCGACCTCGCTACCGCTGC 3429  
Qy 419 CGGTATGTGATATCTGAGTGTGCGAGTGCCTTCAGCGCAGATGAACAGTTTAAG 478  
Db 3428 CGGTGAGCGTATCGCGGCTGCTCGGCAATCCCAAGGATCAGAGAGTTCCGCG 3369  
Qy 479 CATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGGCT 538  
Db 3368 AGTGGTGGACATCATCTCAGCAAGAGGCTGAGTATCCCAACCTCCGAGAGCT 3309  
Qy 539 TTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGCTTTTTCGCGCATCATAG 598  
Db 3308 TCACCGAGACGCTGGGCGCCCGCTCGAGGAGTGTGTCGGAATTCGTGTCGCGCGAGATCG 3249  
Qy 599 AAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGAAGCGAAGAA 658

```
Db 3248 CCCAAGCGCGCGGAAACCGAAGGACGACCTGANTCAGCGCGCCTCTGTGCGCGGAGGTG 3189
Qy 659 CAGGCGAGAAAGCTGTCCGGTGAAGAGCTGATTCGGTTCGACGCTGTCTGTCTGTGGCG 718
Db 3188 ACGGGCGAAGCTGACCGACGAGGAAGTGTCAACATCGTCCGGCTGTCTCACCGCGCG 3129
Qy 719 GAAATGAACCACTCAAAACCTGATTTCAATGCGGATGTACAGCATATTAGAAAGCCAG 778
Db 3128 GGCACATCTCCAGCGCGACGCTCTCAGCAACCTGTTCTCTGTGTCTGGAGGAGCACCGC 3069
Qy 779 GGGTTTACAGGAAGTCCGACGCACTCTGAATGATGCTCAGGCAAGTGGAGGAGCCT 838
Db 3068 AGGACAGGCGCGGTCCGCGCGACCGAGCCTCTGTCCGGCGGTGATCGAGGAGACGC 3009
Qy 839 TCGGTTTCAGAGCGCGCGCCCGCGTTCAGGCGCATTCGCAAGCGGATACGAGATCG 898
Db 3008 TCGGCTACCGTCCCGGTTCACTGATCTCTCGGATCTCTGACGAGGACACCGCATCC 2949
Qy 899 GGGGCGACCTGATTAAGAGGTGATATGGTTTGGCGTTTGGGATCGGCAAAATCGTG 958
Db 2948 TCGGCGACCCCATGCGCAAGGCGCAGATGATGATCGCTGTGATCGCTCCGGAAACCGC 2889
Qy 959 ATGAAGCAAGTTTGCACGACGCACTCTTGCATATCCGCGCCATCCCAATCCGCATA 1018
Db 2888 ACACGAGGTGTTCAGGACCGGACACCTTCGACATCCGACGCGAGTCAACAGCAC 2829
Qy 1019 TTGCGTTTGGCCACGCGCATCCATTTTTCGTTTGGGCGCGCGTTCGCGCTTTTGAAGCAA 1078
Db 2828 TGGCGTTCGCGCAACGCGCATCCACTGCTTGGCGCGTTCCTGCGCAGGCTGAGGCGA 2769
Qy 1079 ATATC 1083
Db 2768 AGGTC 2764

RESULT 5
US-10-156-761-576
; Sequence 576, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 576
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1197)
US-10-156-761-576

Query Match 11.4%; Score 135.4; DB 15; Length 1197;
Best Local Similarity 48.9%; Pred. No. 1.7e-32;
Matches 364; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 344 TCACAGATGAATGATTCAAAAATTTACGGGCGGAGTGAGTTGACCTTGTTCAGATT 403
Db 347 TGACCAAGGAAGTGGCGGAAGTGTCCAGAGGGCGAGATCGACGTCGTGACGACT 406
```

```
Qy 404 TTTCTATACCGGCTTCGGTTATTGTGATATCTGAGCTGTCTGGAGTGCCTTTCAGCGAGA 453
Db 407 TCGCTTACCACTTCCCGGTGACCGGTGATCTGCGGCTGTCTCGGCATCCCGGACAGGACG 456
Qy 464 TGGACAGTTTAAAGCATGTCGTATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAG 523
Db 467 AGCAGCTCTTCAGGACTGGACCGATACCTCTGCTCGCTCCGCGGACATCGGGCCCCGAG 526
Qy 524 AAGCTGAAAAGCTTTTTCGAAAGACGAGATAAGTGTGAGGAAGAACTGGCGCGGTTTT 583
Db 527 GTGACACCCCGGAAACCGGACCGAGCGCGCGACCGAGCGCAGCAGGAGATGGGCGAGTACC 586
Qy 584 TTGCGCGCATCATAGAAAGAAAGCGAAACAAACCGGAAACAGGATATATTTTATTTTAG 643
Db 587 TGTGTCAACTCGCCGAAACAGCGCGCGGTTCGGGCCACCGCGGACATGCTCTCCGACCTCG 646
Qy 644 TGGACGCGAAGAAACAGCGGAGAGCTCTCGGTGAAGAGCTGATTCGTTTTTGCACGC 703
Db 647 TCAACGAAACCGGACCGCGCGCAGCGCTCAGCGAGGAGGACCTGGCGGCGGAAACCATCC 706
Qy 704 TGTGCTGTGTGCGCGGAAATGAACCACTACAAACCTGATTTTCAAATGCGATCTACAGCA 763
Db 707 TGTGTTTCATCGCGGACACGAGACCAACCGTCAATCTGATCGCAACGCGGCTCCTCACCC 766
Qy 764 TATTAGAAACGCGCAGCGGTTTACGAGAACTGCGCAGGCCATCTGAACTGATGCTCAGG 823
Db 767 TGTGCGCGCGCGCGCAACTGGAACCGTCTGCGCGAAGACCCGACCTGCTGCGCGGAG 826
Qy 824 CAGTGAGGAGAGCTTTCGCTTTTCAGAGCGCGCGCGCGGTTTTGAGCGGCACTTGCACAGC 883
Db 827 CGGTGAGGAACTGTCTGCGCTACGAGCCCCCGTCCACATGCGCGAGCGGTTCCCTCG 886
Qy 884 GCGATACGAGATCGGCGGCGACCTGATTAAGAAAGTGTATGTTTGGCGTTTGTGG 943
Db 887 TCGACATCGAAGTCCGCGGACCAACGATCCCGGTGCGACGTCGCTCATCTGCGCGTGG 946
Qy 944 CATCGCAATCTGTGATGAAGCAAGTTTGAACGAGCGGACATGTTTGTATTCGCGCGCC 1003
Db 947 CCTCGCGAGCGCGGACCGGATGCGGTTTCAGCGAACCAGCCGTTTCGACCCACCGGCC 1006
Qy 1004 ATCCCAATCGCATATTCGCTTTGCGCAGCGGATCCATTTTTCCTTGGGCCCCCGTTG 1063
Db 1007 CGGACAAACAGCAGCTCGGCTTCGGCAGCGGATATCCACTGTCTTCGGCGGCCACTCG 1066
Qy 1064 CCGCTCTTGAAGCAATATCGGCTT 1088
Db 1067 CCGTATCGAAGCGGAGCGCGCT 1091
```

## RESULT 6

```
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
```

```

; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      11.4%; Score 135.4; DB 15; Length 9025608;
Best Local Similarity 48.9%; Pred. No. 9.1e-30;
Matches 364; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

QY 344 TCACAGTGAAGTCAAAATTTTCAAGGCGCAGTGAATTTGACCTTGTTCACGATTT 403
DB 737658 TGACCAAGGAAGTGGCGGAAGTGTTCAGGAGGCGAGCATCGAGCTGTCGACGACT 737717

QY 404 TTTTCATACCCGCTTCCGGTTATTCTGTATATCTGAGTGTCTGGAGTGCCTTCAGCGAGA 463
DB 737718 TCGGCTACCCACTGCGCGTGNACCGTGTCTGCCCTGTCTGGCATCCGACAGGACG 737777

QY 464 TGAACAGATTAAAGCATGCTGATCTTCTGTGTCAGTACACGGAAGATAAAGTGAAG 523
DB 737778 AGCAGCTTTCCAGGACTGGACCGGATACCTCTGCGTCCGCGACATCGGCGCCGAAG 737837

QY 524 AAGCTGAAAGAGCTTTTTCGAAGAACGAGATATGTGAGGAAGACTGCGCGGTTT 583
DB 737838 GTGACACCGCGAAGCGGACCGAGCGCGCGGCGGAGGAGATGGCGGAGTACC 737897

QY 584 TTGCGCGCATCATAGAGAAAGGCGAAACCGGAACAGGATATTTCTATTATTAG 643
DB 737898 TGTGTCAACTGCGGACAGCGCGCGGTGCGGCCACCGGGGACATGCTCTCCGACCTCG 737957

QY 644 TGAAGCGGAAGAAACAGGCGAGAGCTGTCCGCTGAGAGCTGATTCGCTTTTGGACGC 703
DB 737958 TCAACGAACCGGACCGCGCGGCGACGGCTCAGCAGGAGGAGCTGCGCGCGAACCACTCC 738017

QY 704 TGCTGCTGGTGGCGGGAATGAAACCACTACAACTGATTTCAAAATGCGATGTACAGCA 763
DB 738018 TGCTGTTCAATCGCGGACAGACGAGACCGGTCATCTGATGCGCAACGGCTCTCACCC 738077

QY 764 TATTAGAAAGCGGAGCGGTTTACGAGGAACCTGCGAGCAATCTCTGAATGATGCTCAGG 823
DB 738078 TGTGTCGCGCGCGGACCACTGGAACCGTCTGCGGGAAGACCCGAGCTGTCGCGGAG 738137

QY 824 CAGTGAAGGAAGCTTTCGCTTTAGAGCGCGCGCGCGGTTCGAGCGGATTCGCAAGC 883
DB 738138 CGGTGAGGAAGCTGCTGCGCTACAGCGCGCGGTCGCAATGCGCGAGCGGTTCCCTCG 738197

QY 884 GGGATACGAGATCGGGGCGCACTGATTAAGAAAGTGATGTTTGGCGTTTGTGG 943
DB 738198 TCGACATCGAGCTGCGCGGACCAAGATCCCGGTGSCAGCTCGCTATCTGCGGCTGG 738257

QY 944 CATCGGCAATCTGTATGAAGCAAAAGTTTGAACGCGACATGTTTGTATATCCGCGGCC 1003
DB 738258 CCTCGGCGAGCGCGACCGCGATGCGGTTTCAAGAAACCGGATTCGACCCACCGCC 738317

QY 1004 ATCCCATCGCATATGCGTTTGGCCACGACATCTTTTGGGCTTGGGCGCGCGCTTG 1063
DB 738318 CGGACAAACGACAGCTGCGGCTTGGCAGCGGTATCCACCTGTGCTTGGCGCGCACTCG 738377

QY 1064 CCGCTCTTGAAGCAAAATATCGCGTT 1088
DB 738378 CCGGTATCGAAGCGGAGCGCGGCT 738402

RESULT 7
US-10-229-148B-1/c
; Sequence 1, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
```

```
? NAME/KEY: CDS
? LOCATION: Complement((19063)..(20229))
? OTHER INFORMATION: ORF26
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((120307)..(21743))
? OTHER INFORMATION: ORF25
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((21733)..(22527))
? OTHER INFORMATION: ORF24
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((22534)..(23571))
? OTHER INFORMATION: ORF23
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((23555)..(24463))
? OTHER INFORMATION: ORF22
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((24460)..(25650))
? OTHER INFORMATION: ORF21
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((25647)..(26105))
? OTHER INFORMATION: ORF20
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((26180)..(27391))
? OTHER INFORMATION: ORF19
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (27937)..(28983)
? OTHER INFORMATION: ORF18
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (29244)..(42779)
? OTHER INFORMATION: ORF1
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (42823)..(48657)
? OTHER INFORMATION: ORF2
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (48712)..(59802)
? OTHER INFORMATION: ORF3
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (59850)..(64556)
? OTHER INFORMATION: ORF4
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (64687)..(70365)
? OTHER INFORMATION: ORF5
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (70365)..(71078)
? OTHER INFORMATION: ORF6
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (71113)..(72360)
? OTHER INFORMATION: ORF7
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (72400)..(73665)
? OTHER INFORMATION: ORF8
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (73694)..(75043)
? OTHER INFORMATION: ORF9
? FEATURE:
```

```
? NAME/KEY: CDS
? LOCATION: Complement((75899)..(76570))
? OTHER INFORMATION: ORF10
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((76602)..(77765))
? OTHER INFORMATION: ORF11
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (78039)..(79313)
? OTHER INFORMATION: ORF12
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((79391)..(81052))
? OTHER INFORMATION: ORF13
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (81541)..(82356)
? OTHER INFORMATION: ORF14
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (82760)..(83362)
? OTHER INFORMATION: ORF15
? FEATURE:
? NAME/KEY: CDS
? LOCATION: Complement((83495)..(84142))
? OTHER INFORMATION: ORF16
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (84329)..(84428)
? OTHER INFORMATION: ORF17 (fragment)
US-10-229-1488-1
```

```
Query Match 10.2%; Score 121; DB 17; Length 84428;
Best Local Similarity 47.2%; Pred No. 1.6e-26;
Matches 401; Conservative 0; Mismatches 445; Indels 3; Gaps 2;

QY 239 CCATCATTAACATGGACCCGCGAAGCATACAAAATCCGTTACGTCTGTGAACAAGCCT 298
DB 27147 CCATCGGCGCCCTCGACCCGCCGCCACGGCCCGATCGCAAGCTGTCTACGCCAGGCGT 27088

QY 299 TTACTCCGCGGTGATGAAGCATGGACCGAGATTTCAAGAAATCAGATGAATGA 358
DB 27087 TCACCCCGCGCGGATGGCCGCTCTGGAACCCCGCATCCGGSCCCTCCTCAGGAGCTCC 27028

QY 359 TTCAAAAATTCAGGGCGCGAGTGTGACCTTTGTTACGATTTTTCATACCCGCTTC 418
DB 27027 TCGATCGGTGGCGGCCAGGAGACCATCGAGTGTGGGACCTCTCTACGCCCTGC 26968

QY 419 CGGTTATTGTGATATCTGAGCTGTGGGAGTGCCTTCAGCGCAGATGGAAAGTTAAAG 478
DB 26967 CGGTATCTGATCTTCTGTGTCTAGTACCCGCGATACCGTCCGGTACCGCGATGTGTTCCGG 26908

QY 479 CATGCTCTGATCTTCTGTGTCTAGTACCCGAGGATAAAGTCAAGAGCTGAAAAGCCT 538
DB 26907 GGTGGGTGACACCTCTGCTCACCAAGAGGGCCTGGAGTACCCGAACCTCCCGGACAACT 26848

QY 539 TTTTGGAAAGACGAGATAAGTGTGAGGAAGAACTGGCCGCGTTTTTTTGGCGCATCATAG 598
DB 26847 TCAGCGAGACGATCGCCCGCCCTCAAGSAGATGACCGACTATCTCTGACCCAGATCC 26788

QY 599 AAGAAAGCGAAACAAACCGGAACAGGATATTATTTCTATTTTAGTGGAGCGGAGAAA 658
DB 26787 ACGCCAAAGCGGAGGCCCGGTGCGACGACCTGATCAGCGGCCCTGTGTCAGGCGCGAGCAG 26728

QY 659 CAGCGAGAGCTGTCCGCTGAAGAGCTGATCCCGTTTGCAACGCTGCTGCTGGTGGCG 718
DB 26727 ACGGCCGCAAGCTCACCGAGTGGAGATGCTCAACATCGTGGCCCTGCTGCTCAACCGCG 26669

QY 719 GAAATGAAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGAAAAGCCAG 778
DB 26667 GCCACGTCTCGTCAAGCAGCGTGTCTCAGCAACCTCTTCTGTGTACTGGAGGAGACCCGC 26608
```

QY 779 GCGTTTACGAGAACTCGCAGCCCATCTGAATGATGCTCTCAGCAGTGGAGAAAGCCT 838  
Db 26607 AGCGCTGGCGGACCTCGCGCCGACCGCGAGCTGGTGAACCGTGGAGGAGACAC 26548  
QY 839 TCGTTTACAGAGCCCGCCCGCTTTGAGCGCATTCGCAAGCGGATACGAGATCG 898  
Db 26547 TCGGCTACCGAGCGCCCTTCAACAACATCTTCCGCTTCTCAAGGAGACACCGACATCC 26488  
QY 899 GGGGCACTTGATTAAGAAGGTGATAGTGTGTTGGGCTTTGTGSCATCGGCAATCGTG 958  
Db 26487 TCGCCCGGGAATGAAGAGGCGCAGATGTCATCGCTCGAGCCAGTCGGACCGCG 26428  
QY 959 ATGAAGCAAGTTTGAAGACCGCAGCATGTTTGTATATCGCGCCGATCCCA---ATCGCG 1015  
Db 26427 ACCCGGAACATCTCCCGAGCGCGACACCTTCGACATCCGACGCTCAAGCAGCTCCCGCC 26368  
QY 1016 ATATTGGTTTGGCCAGCGCATTCATTTTCCCTTGGGCGCCGCTTGCCTGCTTGAAG 1075  
Db 26367 ACATGGCGTTTCGGCATCGGTATTCACACTGCTGCGCGCTTCTTGACAGTCAGGAAG 26308  
QY 1076 CAATATCG 1084  
Db 26307 GCAAGTGG 26299

## RESULT 8

US-10-214-446-39  
; Sequence 39, Application US/10214446  
; Publication No. US20030180742A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David  
; APPLICANT: Burk, Mark J.  
; APPLICANT: Hitchman, Tim  
; APPLICANT: Pujol, Catherine  
; APPLICANT: Richardson, Toby  
; APPLICANT: Short, Jay M.  
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING  
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM  
; FILE REFERENCE: 09010-500001  
; CURRENT APPLICATION NUMBER: US/10/214,446  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/309,497  
; PRIOR FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 1215  
; TYPE: DNA  
; ORGANISM: Bacterial  
US-10-214-446-39

Query Match 10.1%; Score 120.4; DB 15; Length 1215;  
Best Local Similarity 49.6%; Pred. No. 1.3e-27;  
Matches 418; Conservative 0; Mismatches 391; Indels 33; Gaps 3;  
QY 241 ATCATTAACATGACCGCCGCGAAGCATACAAAATCCGTTGCTGCTGAAACAAAGCCTTT 300  
Db 277 ATCTCAGCATGACCGCCGCGGACACACCCCGCTTGGCCAGCGCTTC 336  
QY 301 ACTCCGCGGTGTAAGCATGGACCGGAGATTCAAGAAATCACAGATGAAGTATT 360  
Db 337 ACCATGCAACAGGTGGAGAAAGTTGCGCCCGCGGTGGGAGCTGGCCGAGAGCTGATC 396  
QY 361 CAAAAATTTTCAGGGCGCGAGTGA---GTTTGACCTTGTTCAGATTTTTCATACCGCTT 417  
Db 397 GACAAGATGTTCCGACCGCGCCCGCTGCACTGGTCGAGAGTTTCGCGCTGCGGTG 456  
QY 418 CCGGTTATGTGATATCTGAGCTGCTGGAGAGTGCCTTCAGCGCAGATGGAACAGTTTAA 477  
Db 457 CCGGTCGGGGTGTATCTGCACTGTCTGCGCTGCGGTGCGGACCGCTCCGCGCTTCCGG 516  
QY 478 GCATGTTCTGATCTTCTGCTCAGTACACCGGAGGATTAAGTGAAGAGCTGAAAAAGCC 537

Db 517 GCGTGGAGCGAGCGCGCTGTCTCCACAGTTCCCTGACGCGCGAGGAGTTGACGCCAAC 576  
QY 538 TTTTGGAGAAACAGAGATAAGTGTGAGGAAGAACTGGCCGGTTTTTTTCGCGCATCAT 597  
Db 577 C-----AGGAGAACTGCGGGCTACATCGGGGGTTGATC 612  
QY 598 GAAGAAAGCGAAACAAACCGGAACAGGATATTTCTATTTAGTGAAGCGGAAGAA 657  
Db 613 GAGGATCACCGGGCGCTCCGCTGAGGACCTGATCACCGGCTGATCGAGCCCGGAC 672  
QY 658 ACAGCGAGAGCTGCTCGGTGAGAGAGCTGATCCGTTTTCACGCTGCTGCTGCTGCTG 717  
Db 673 CGCGACCGCGCTGACCGAGCAGAGTTGGTGGACCTGTGGTGGGATCTCTGTTGCC 732  
QY 718 GGAATGAAACCACTACAAACCTGATTTTCAAAATGCGATGTACAGCATATTAGAAACGCC 777  
Db 733 GGCACAGAGACCAACCGCCACGATCCCACTTCGTTGGTGAAGCTGCTGSAACCGGCC 792  
QY 778 GCGTTTACGAGGAACTGCGGAGCGCATCTGAACTGATGCTCAGGAGTGAAGAGGCC 837  
Db 793 GAGCAGTGAACCCGCTGCGGAGGACCCGAGCTGGTCCCGACCGCGGTTCAGGAGCTG 852  
QY 838 TTGCGTTTACAGCGCGCGCCCGCTT-----TTGAGCGCATTCGCAAGCGGATAGC 891  
Db 853 ATGGCTTTCGCTGGCGAGCGCTGCTGCTCCCGGTACGCCACCGAGGACGTG 912  
QY 892 GAGATCGGGGGCACCTGATTAAGAAAGTGTGATATGTTTGGCGTGTGTCATC39CA 951  
Db 913 GAGTTCGCGCGCACGCTGCTGCGCGCGGAGCGCGTGTGCTGGTGGCGTCCGGGGCGCC 972  
QY 952 AATGCTGATGAAGCAAGTTTGAAGAGCGGACATGTTTGAATATCGCGCGCATCCGAT 1011  
Db 973 AACCAGACCGCGCGAGTTTCAGCGCGCGAGGAGCTGGACCTGCGCGCGGAGGCGCAC 1032  
QY 1012 CGCATATTGCTTGGGCGACCGCATTCATTTTCCCTTGGGCGCCCGCTTCCCGCTCTT 1071  
Db 1033 CAGACCTCGGTTTCGGCATGGGCTCCACCACTGCTCGGGGCGCGCTGCGCGCGCTG 1092  
QY 1072 GA 1073  
Db 1093 GA 1094

## RESULT 9

US-10-156-761-2364  
; Sequence 2364, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2364  
; LENGTH: 1197  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1197)  
US-10-156-761-2364

Query Match 9.9%; Score 117.8; DB 15; Length 1197;  
Best Local Similarity 47.9%; Pred. No. 8.8e-27;  
Matches 406; Conservative 0; Mismatches 432; Indels 9; Gaps 2;

QY 245 TTAACATGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 304  
DB TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 304  
QY 239 TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 298  
DB TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 298  
QY 305 CGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
DB CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
QY 299 CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 358  
DB CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 358  
QY 365 AATTTCAGGCGCGAGTGTGAGTGTGACCTTGTTCAGATTTTTCATACCGCTTCGGTTA 424  
DB CGATCGTCCCGCGGAGCGAGTGTGAGTGTGACCTTGTTCAGATTTTTCATACCG 418  
QY 425 TTGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 484  
DB TTGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 478  
QY 419 CGGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 478  
DB CGGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 478  
QY 485 CTGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544  
DB CTGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 532  
QY 479 CGAA-----GACCTTCGGCGGTGAGCAACCGACTCCCGCGGCGGCGGTG 532  
DB CGAA-----GACCTTCGGCGGTGAGCAACCGACTCCCGCGGCGGCGGTG 532  
QY 545 AAGAACGAGATTAAGTGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604  
DB AAGAACGAGATTAAGTGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604  
QY 533 ACGCGCGCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 592  
DB ACGCGCGCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 592  
QY 605 AGCGAAACAAACCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 564  
DB AGCGAAACAAACCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 564  
QY 593 GCGGTGAGGAGCGGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 552  
DB GCGGTGAGGAGCGGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 552  
QY 665 AGAAGCTGTGCGGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 724  
DB AGAAGCTGTGCGGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 724  
QY 653 ACCGCTGAAACGAGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 712  
DB ACCGCTGAAACGAGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 712  
QY 725 AAACCACTACAAACCTGATTTTCAATGAGATGATGATGATGATGATGATGATGATGAT 784  
DB AAACCACTACAAACCTGATTTTCAATGAGATGATGATGATGATGATGATGATGATGAT 772  
QY 713 AGACGACGATCAATCTGATCGGCAACGCGACCTGCGCGCTGCTGCGCGACCCCGACCAAC 772  
DB AGACGACGATCAATCTGATCGGCAACGCGACCTGCGCGCTGCTGCGCGACCCCGACCAAC 772  
QY 785 ACGAGGAATCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 844  
DB ACGAGGAATCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 844  
QY 773 TCGGCTGTGCGGAGGAGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 832  
DB TCGGCTGTGCGGAGGAGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 832  
QY 845 TCAGAGCGCGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 901  
DB TCAGAGCGCGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 901  
QY 833 ACGAGCGGTGATGAGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 892  
DB ACGAGCGGTGATGAGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 892  
QY 902 GGCACCTGATTAAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 961  
DB GGCACCTGATTAAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 952  
QY 893 GCAAGCGCATGAGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 952  
DB GCAAGCGCATGAGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 952  
QY 962 AAGCAAGTGTGACAGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1021  
DB AAGCAAGTGTGACAGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1021  
QY 953 GGGCGAGTGTGCGGAGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1012  
DB GGGCGAGTGTGCGGAGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1012  
QY 1022 CGTTTGGCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1081  
DB CGTTTGGCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1072  
QY 1013 CTTTGGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1072  
DB CTTTGGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1072  
QY 1082 TCGGTT 1088  
DB TCGGTT 1079

## RESULT 10

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN

Query Match 9.9%; Score 117.8; DB 15; Length 9025608;  
Best Local Similarity 47.9%; Pred. No. 4.8e-24;  
Matches 406; Conservative 0; Mismatches 432; Indels 9; Gaps 2;

QY 245 TTAACATGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 304  
DB TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 304  
QY 305 CGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
DB CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
QY 2902518 TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 2902519  
DB TGACACGAGCCGCGGAGCATACAAATAATCGTTTACGTGTAACAAAGCCTTTACTC 2902519  
QY 2902518 CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
DB CCGCGGTGATGAGCAATGGGACCGAGATTAAGAAATACAGATGAATGATTCATAA 364  
QY 365 AATTTCAGGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 424  
DB AATTTCAGGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 424  
QY 2902458 CGATGCTGCCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902399  
DB CGATGCTGCCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902399  
QY 425 TTGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 484  
DB TTGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 484  
QY 2902398 CGGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902339  
DB CGGTGATCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902339  
QY 485 CTGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544  
DB CTGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544  
QY 2902338 CGAA-----GACCTTCGGCGGTGAGCAACCGACTCCCGCGGCGGCGGTG 2902285  
DB CGAA-----GACCTTCGGCGGTGAGCAACCGACTCCCGCGGCGGCGGTG 2902285  
QY 545 AAGAACGAGATTAAGTGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604  
DB AAGAACGAGATTAAGTGTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604  
QY 2902284 ACGCGCGGCGGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902225  
DB ACGCGCGGCGGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902225  
QY 605 ACGGAAACAAACCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 664  
DB ACGGAAACAAACCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 664  
QY 2902224 GCGGTGAGGAGCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902165  
DB GCGGTGAGGAGCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902165  
QY 665 AAGAGCTGTGCGGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 724  
DB AAGAGCTGTGCGGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 724  
QY 2902164 ACCGCTGAAACGAGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902105  
DB ACCGCTGAAACGAGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902105  
QY 725 AAACCACTACAAACCTGATTTTCAATGAGATGATGATGATGATGATGATGATGATGATGAT 784  
DB AAACCACTACAAACCTGATTTTCAATGAGATGATGATGATGATGATGATGATGATGATGAT 784  
QY 2902104 AGACGACGATCAATCTGATCGGCAACCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902045  
DB AGACGACGATCAATCTGATCGGCAACCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2902045  
QY 785 ACGAGGAATCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 844  
DB ACGAGGAATCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 844  
QY 2902044 TCGGCTGTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2901985  
DB TCGGCTGTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2901985  
QY 845 TCAGAGCGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 901  
DB TCAGAGCGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 901  
QY 2901984 ACGAGCGGTGCTTCGAGACGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2901925  
DB ACGAGCGGTGCTTCGAGACGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2901925  
QY 902 GGCACCTGATTAAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 961  
DB GGCACCTGATTAAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 961





Db 291 CACCCTGGTCTGCTGCAACAACTCCGGAGCCCGTACGGAGACCGCCACCGGCTTCT 350  
QY 360 TCBAARAAATTCAGGGGCGAGTGTTCACCTTGTTCACGATTTTTCATACCCGCTTCC 419  
Db 351 CAGCCACTCGGGAGCGACCGGAGACCGACTCATCGCTCTGTAGCGCGCCCTGCG 410  
QY 420 GGTATTGTGATCTGAGCTCTGGAGTGTCTTCAGCGCAGATGGAACAGTTTAAAGC 479  
Db 411 GATCACCGTCTATCTGACCTCTCGGCTGCGGACGAAACCGCGGAGCTTCCGGCG 470  
QY 480 ATGGTCTGATCTTCTGCTCAGTACACCGAGAGATTAAGTGAAGAGCTGAAGAGCTT 539  
Db 471 CTGGACCGACCGCTCTGACCCCGGACCGCGCCCGCGGAGCTGCGCAGGAAATCCGT 530  
QY 540 TTGTGAAGAAAGAGATAAGTGTGAGGAAGAACTGGCGCGGTTTTTGTCCGCGCATATAGA 599  
Db 531 CGTCTCGT-----TGCTCGCTTCTTTCACCGGCTCTCTCG 566  
QY 600 AGAAAGGAAACAAACCGGAACAGATATATTTCTATTTTGTGGAAGCGGAAGAAAC 659  
Db 567 CGACAAGCGCAAGAACCCCGCGGACGACCTGCTCTCGACCTCATCGCGCTGCAGGAAGA 626  
QY 660 AGCGAGAGCTGTCCGCTGAAGAGCTGATTCGTTTTTTCAGCTGCTCTGCTGTCGCGG 719  
Db 627 GGGGACCGGCTCACCGAGGACGAGCTGATGTCCTCGCTTCTCTCTCTCTCTCTCTCT 686  
QY 720 AAATGAACCACTACAAACCTGATTTCAATTCGATGTACAGCATATTAGAAACGCGAG 779  
Db 687 CTRGAGAAACCGTGACCTCTTCGGAAACGCGCTACTGGGCTGCTGCGCCATCCGA 746  
QY 780 CGTTTACGAGAACTGCGGAGCCATCTCTGAATGATGCTCAGGAGTGAAGAGAGCTT 839  
Db 747 CGAGCTCCCGCGCTCCCGAGGACCGGACCGCTGCGGACCGCGCTCGGGAGTTCCG 806  
QY 840 CGCTTTCAGAGCGCGCGCCCGGTTT---TGAGGCGCATTCGACGCGGATACGAGAT 896  
Db 807 CGCTTACGAGGCGCGCGCTCTCGCATCCCGCTTCCCGGTGCGGAGCTGACGAT 866  
QY 897 CGGGGCGCACCTGATTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCCGCAATCG 956  
Db 867 CGCGGGGTACCGCTGCGCGCGGGGAGACCGTGTCTGTCTGTCTGCTGCTGCTGCTG 926  
QY 957 TGATGAACAAAGTTTGAACAGCCGACATGTTTATATTCGCGCGCATTCGCAATCCGA 1016  
Db 927 TGACCGGACCGGTTTCCCGGACCCGACCGGCTCGATCTCGGCGCGGACGCGCGGCA 986  
QY 1017 TATTGCTTTGCGCACGCGCATCTTTTTCCTTGGCGCCCGCTTGGCGCTTGTGAAGC 1076  
Db 987 CTTGCGCTCGGCGCGGCGCTTCCACTTACTGCTTGGCGCGCGCTGCGCGCTGAGAC 1046  
QY 1077 AAATATCGC 1085  
Db 1047 GGAGTCCG 1055

## RESULT 13

US-10-214-446-55  
; Sequence 55, Application US/10214446  
; Publication No. US20030180742A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David  
; APPLICANT: Burk, Mark J.  
; APPLICANT: Hitchman, Tim  
; APPLICANT: Pujol, Catherine  
; APPLICANT: Richardson, Toby  
; APPLICANT: Short, Jay M.  
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING  
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM  
; FILE REFERENCE: 09010-500001  
; CURRENT APPLICATION NUMBER: US/10/214,446  
; CURRENT FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/309,497  
; PRIOR FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 55  
; LENGTH: 1248  
; TYPE: DNA  
; ORGANISM: Bacterial  
US-10-214-446-55

Query Match 9.6%; Score 114.2; DB 15; Length 1248;  
Best Local Similarity 46.7%; Pred. No. 1.3e-25;  
Matches 399; Conservative 0; Mismatches 453; Indels 3; Gaps 1;

QY 230 TTGGAATTCATCATTAACATGACCCCGGAGACATCAAAATTCCTTGTGAGTCTGA 289  
Db 263 TCAGAACTGGCTGCTCTTCTTGGACCCCGCCCGGACACGCGGCTGCGCCAGGTGG 322  
QY 290 ACAAGACCTTTTACTCCGCGCTGATGAACATGGAAACCGGAAATTCAGAAATCAG 349  
Db 323 CCCCOCGCTGAGCGTCCGCGCTGACGGGCTTTCGCTCCCGCGTACGGAGATCGCG 382  
QY 350 ATGAACCTGATTCAAAAATTCAGGGGCGCAGTGTGACCTTGTTCACGATTTTTCAT 409  
Db 383 AGGAATCTGACGCGCCCTCGCGCGCGCGCGCTGGTGAATCTGTCGAGGCTTCGCG 442  
QY 410 ACCCGCTTCGCTTATTTGTAATCTGAGTCTGCGAGTGTGCTTCAGCGCAGATGGA 469  
Db 443 CCCCCTTCGCTGCTGCTGCTGCGGGGCTGCTCGGGTTCGACCCCGCTGCGTGGCG 502  
QY 470 AGTTTAAACATGCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAGCTG 529  
Db 503 GGTTCGGGAGAGGCG---CTCGCCCTCCAGCGTTCGCGCGGACCGAGGCGACCGGT 559  
QY 530 AAAAGACCTTTTGGAAAGAACGAGATAAGTGAAGAGAACTGGCGCGCTTTTTCGCG 589  
Db 560 CCGCGCGCGCTCGACCGGCGCGCGCGCGCGCGCGCGCGCTTACGCGCTTTCGCG 619  
QY 590 GATCATGAAGAAAGCGAAACAAACCGGAACAGGATATTTCTATTTTAGTGGAG 649  
Db 620 CCGAGCTGCGCGCGCGCTCGCGAGACCGGGGAGACCTGCTCTCGCGCTGCGCGCG 679  
QY 650 CGGAAGAAACCGCGAGAGAGCTGCGGTGAAGAGCTGATTCGCTTTTCACGCTCTCG 709  
Db 680 CCGCGCGCGAGACCGCTCGCTGGGAGACCGCGCTGACGCTGACCTGCTCCACCTCC 739  
QY 710 TGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACGATATAG 769  
Db 740 TGACCGCGGCGCACGAGACGACGCGGCTCTGCGCAAGCGGCTGCTCGCGCTCTCG 799  
QY 770 AAACGCGAGCGTTTACGAGGAACTGCGCAGCGCTCTGAACTGATGCTCAGGAGTGG 829  
Db 800 CCGCGCGGAGTGGCGGAGAACTCGCGCGGACCGCGGCGCTTTCGGAACGCGCGTGG 859  
QY 830 AGGAAGCTTTCAGAGCGCGCGCGCTTTTGAAGCGCATTTGAGCGGATTCGAAGCGGATA 889  
Db 860 ACAGTTCCTGCGCGCACGACCCACCGTGCAGATGCTACGCGTGGGCGCGCGGAGCG 919  
QY 890 CGGAGATCGGGGCGCACCTGATTTAAAGAGGTGATATGTTTTTGGCGTTTGTGGCATCG 949  
Db 920 CGGAGCTCGCGCGCGAGCGGTCGCGCGGCGGAGACCGGCTCCAGCTGCTGCTGGTTCG 979  
QY 950 GAAATCGTGAAGAAGAAAGTTTGAACAGCCGACATGTTTGAATTCGCGCGCGCATCCCA 1009  
Db 980 CCGACCGGACCCCGCGCGCTTCCGCGACCCCGAGCGGCTGGAACATCCCGCGGACCG 1039  
QY 1010 ATCCGATATTCGCTTGGCGACGCGCATCTTTTTCCTTGGCGCGCGCTTGGCGCTG 1069  
Db 1040 GCGCGCATCGCGCTGCTGCTGGGCTACCTACTGCTGGGCGCGCGCTGGCGCGCG 1099  
QY 1070 TTGAAGCAAAATATCG 1084  
Db 1100 CGGAGCGGAGATCG 1114

## RESULT 14

```
US-09-738-626-617
; Sequence 617, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 617
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-617

Query Match          9.5%; Score 112.8; DB 9; Length 1290;
Best Local Similarity 54.6%; Pred. No. 3.9e-25;
Matches 247; Conservative 0; Mismatches 202; Indels 3; Gaps 1;

Qy 635 CTATTTAGTGAAGCGGAAGAAACAGCGGAGAGCTGTCGGTGAAGAGCTGATCCGT 694
Db 719 CTGATCTAGTGGGACGACAGCAAGAGGGTCAAGAAATCACCGATCATGAGATTGCTCT 778

Qy 695 TTTCACGCTGCTGCTGTCGTCGCGGAAATGAACACCTACAACTGATTTCAATGCGA 754
Db 779 TGCTGTATCTCCTGCTCTTTTGGGGGACGAAACACCACTGATTCCTCAATGTT 838

Qy 755 TGTACAGCATATTAGAAACGCGCGCTTTACAGGAACTCGCGAGCCATCTGAACTGA 814
Db 839 TCCGAGTTCCTCGATCATCACAGCAGTGGCAAGCCATCTAGAGAAATCCAAACTGA 898

Qy 815 TGCCTCAGGAGTGGAGGAGCCCTTGCTTTTCAGAGCGCGCCCGCTTTGAGGCGCA 874
Db 899 TTCTGCGCAGTGGATGAGTCTTTGCGGTACTCGGCTCGATCGTGGGGTGGGCTGAA 958

Qy 875 TTGCCAAGCGGGATACGAGATCGGGGGGACCTGATTAAAGAGGATGATATGTTTGG 934
Db 959 AAGCATTAAGACACCGAGATCGCGCGCTTGCCATTAAAGAGGCGATGCTGTTCTGC 1018

Qy 935 GGTTTGTGGCATCGCAAAATCGTATGAAGCAAAAGTTTGACAGACCGGCACATGTTGATA 994
Db 1019 TGCTCATGGGTTCCGGAACCGCATGAAGCTGCTTTGAAATGGCGAGGAATTCGATA 1078

Qy 995 TCCGCGCG---CATCCCAATCCGATATGCGCTTTGGCCACCGCATCCATTTTGCCTTG 1051
Db 1079 TCAGCGCGCTTAATGCGCGGAGCACCTGCTCTTTGGTTTCGGCATCCACTATTGCTTAG 1138

Qy 1052 GGGCGCGCTTGGCGCTCTTGAAGCAAAATC 1083
Db 1139 GAAACATGCTGCCAAACTTCAAGCAAGATC 1170
```

## RESULT 15

```
US-09-738-626-1/c
; Sequence 1, Application US/09738626

Query Match          9.5%; Score 112.8; DB 9; Length 3309400;
Best Local Similarity 54.6%; Pred. No. 1e-22;
Matches 247; Conservative 0; Mismatches 202; Indels 3; Gaps 1;

Qy 635 CTATTTAGTGAAGCGGAAGAAACAGCGGAGAGCTGTCGGTGAAGAGCTGATCCGT 694
Db 567370 CTGATCTAGTGGGACGACAGCAAGAGGGTCAAGAAATCACCGATCATGAGATTGCTCT 567311

Qy 695 TTTCACGCTGCTGCTGTCGTCGCGGAAATGAACACCTACAACTGATTTCAATGCGA 754
Db 567310 TGCTGTATCTCCTGCTCTTTTGGGGGACGAAACCACTGATTCCTCAATGTT 567251

Qy 755 TGTACAGCATATTAGAAACGCGCGCTTTACAGGAACTGCGAGCCATCTGAACTGA 814
Db 567250 TCCGAGTTCCTCGATCATCACAGCAGTGGCAAGCCATCTAGAGAAATCCAAACTGA 567191

Qy 815 TGCCTCAGGAGTGGAGGAGCCCTTGCTTTTCAGAGCGCGCCCGCTTTGAGGCGCA 874
Db 567190 TTCTGCGGAGTGGATGAGTCTTTGCGGTACTCGGCTCGATCGTGGGGTGGGCTGAA 567131

Qy 875 TTGCCAAGCGGGATACGAGATCGGGGGGACCTGATTAAAGAGGATGATATGTTTGG 934
Db 567130 AAGCATTAAGACACCGAGATCGCGCGCTTGCCATTAAAGAGGCGATGCTGTTCTGC 567071

Qy 935 GGTTTGTGGCATCGCAAAATCGTATGAAGCAAAAGTTTGACAGACCGGCACATGTTGATA 994
Db 567070 TGCTCATGGGTTCCGGAACCGCATGAAGCTCGCTTTGAAATGGCGAGGAATTCGATA 567011

Qy 995 TCCGCGCG---CATCCCAATCCGATATGCGCTTTGGCCACCGCATCCATTTTGCCTTG 1051
Db 567010 TCAGCGCGCTTAATGCGCGGAGCACCTGCTCTTTGGTTTCGGCATCCACTATTGCTTAG 566951

Qy 1052 GGGCGCGCTTGGCGCTCTTGAAGCAAAATC 1083
Db 566950 GAAACATGCTGCCAAACTTCAAGCAAGATC 566919
```

Search completed: May 28, 2004, 23:12:11  
Job time : 468 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 20:10:15 ; Search time 82 Seconds  
(without alignments)  
8060.323 Million cell updates/sec

Title: US-09-869-334B-2  
Perfect score: 1191  
Sequence: 1 atgaatgtttaaacgccg.....gcttcggtgaaatgtaa 1191

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	139.2	11.7	8478	3	US-08-676-818-1
2	139.2	11.7	8478	4	US-09-407-549-1
C 3	118.2	9.9	4403765	3	US-09-103-840A-2
C 4	118.2	9.9	4411529	3	US-09-103-840A-1
5	116.4	9.8	4403765	3	US-09-103-840A-2
6	116.4	9.8	4411529	3	US-09-103-840A-1
7	111.4	9.4	1194	4	US-08-765-907A-9
8	111.4	9.4	4496	4	US-08-765-907A-6
9	109.8	9.2	1314	4	US-09-252-991A-1265
C 10	109.8	9.2	1482	4	US-09-252-991A-1338
11	109.8	9.2	1671	4	US-09-252-991A-1222
12	103.6	8.7	6085	3	US-09-029-603-4
13	88.8	7.5	1233	4	US-09-266-965-24
14	88.8	7.5	12249	4	US-09-266-965-74
15	88.8	7.5	18331	4	US-09-266-965-96
16	87.6	7.4	71989	4	US-09-443-501A-2
C 17	86.8	7.3	536165	4	US-09-214-808-1
18	85.8	7.2	1221	6	5212296-16
19	85.9	7.2	1879	6	5212296-5
20	84.4	7.1	1224	4	US-09-266-965-22
21	84.4	7.1	1227	3	US-09-385-028-23
22	84.4	7.1	1227	4	US-09-726-614-23
23	84.4	7.1	1227	4	US-09-385-040-23
24	84.4	7.1	15079	3	US-09-385-028-1
25	84.4	7.1	15079	4	US-09-726-614-1
26	84.4	7.1	15120	4	US-09-385-040-1
27	84.2	7.1	68750	3	US-09-568-102-1

28	84.2	7.1	68750	4	US-09-568-102-1
29	84.2	7.1	68750	4	US-09-567-969-1
30	84.2	7.1	68750	4	US-09-568-480-1
31	84.2	7.1	68750	4	US-09-568-486-1
32	84.2	7.1	68750	4	US-09-568-472-1
33	84.2	7.1	68750	4	US-09-567-899-1
34	81.2	6.8	1407	4	US-09-252-991A-15866
C 35	81.2	6.8	1731	4	US-09-252-991A-15928
C 36	81.2	6.8	2502	4	US-09-252-991A-15767
37	80.4	6.8	47981	4	US-09-679-279-1
38	72.2	6.1	1251	3	US-09-105-537-38
39	72.2	6.1	5970	3	US-09-320-878-21
40	72.2	6.1	5970	4	US-09-141-908-11
41	72.2	6.1	5970	4	US-09-657-440-21
42	71.2	6.0	50937	3	US-09-428-517-1
43	68	5.7	1735	1	US-08-102-863-10
44	68	5.7	1735	5	PCT-US92-10885-10
45	65	5.5	1242	3	US-09-068-132-1

## ALIGNMENTS

RESULT 1  
US-08-676-818-1  
; Sequence 1, Application US/08676818  
; Patent No. 6057136  
; GENERAL INFORMATION:  
; APPLICANT: Bewer, Stanley Grant  
; APPLICANT: Perkins, John B.  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Pero, Janice G.  
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS  
; TITLE OF INVENTION: SUSTILIS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676,818  
; FILING DATE: 08-JUL-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/239,430  
; FILING DATE: May 6, 1994  
; APPLICATION NUMBER: 08/084,709  
; FILING DATE: June 25, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Freeman, John W.  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 04599/004001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8478  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-676-818-1

Query Match 11.7%; Score 139.2; DB 3; Length 8478;







```
QY 729 CACTACAAACCTGATTTCAATGGGATGTACAGATATTAGTAAACCCGAGGCGTTTACGA 788
Db 872199 CACCCTTAATCTCAATGCGCGGCGTCTGGCGCTGGCGGAGAACCTTGACCACTGCA 872258
QY 789 GGAACTGGCGAGCCATCTGAACTGATGCTCAGGAGTGGAGGAGACCTTGGTTTCAG 848
Db 872259 AACGTGGGAAGCGATTTTGAAGTTGTGCGGATCGATCGAAGAGATCGTGAGGTGGAC 872318
QY 849 AGCGCGCGCCCGGTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGGGGCACT 908
Db 872319 GTCCGCTGCTCACCATCGAAGCGGCGACGCGCTCCGCGGTGAGCTGGCGGCGCGGCC 872378
QY 909 GATTAAGAGGTGATATGTTTGGCGTTTGGCGATCGGCAAAATCGTGATGAAGCAAA 968
Db 872379 GATCGAGCGGCTCAGAAAGTTGTGCTGTGGAGGCGCTCGGCCAACCGTGATCCAGCGT 872438
QY 969 GTTTGACAGACCGCATCATGTTTGATATCGCGCGCATCCCAATCCGATATTCGCTTGG 1028
Db 872439 GTTCAGCGCGGCTCAGAAAGTTGTGCTGTGGAGGCGCTCGGCCAACCGTGATCCAGCGT 872498
QY 1029 CCAGCGCATCCATTTTTCGCTTGGGCGCCCGCTTGCCTGCTTGA 1074
Db 872499 TCAGGGGCTGCACTATTGCTGGCGGCCAATCTGGCTCGGCTGGA 872544

RESULT 7
US-08-765-907A-9
; Sequence 9, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-9

Query Match 9.4%; Score 111.4; DB 4; Length 1194;
Best Local Similarity 48.5%; Pred. No. 2.2e-26;
Matches 415; Conservative 0; Mismatches 416; Indels 24; Gaps 3;

QY 249 CATGGACCCGCGGAGCATACAAAATCCGTTCACTGCTGGAACAAAGCTTTACTCCGCG 308
Db 228 CATGACCCGCGGATGACCCGACCCCTGGCGGCTGTGACCGAGGCTTACACCCCGCG 287
QY 309 CGTGATGAAGCAATGGGACCGAGAAATTCAGAAATCAAGATGAATGATTCAGAAATTT 368
Db 288 CACCGTCCCGACCTCGAACCAGCGGTCCACGAACTGGCGCGGCACTGCTCGACGCGGT 347
QY 369 TCAGGGGCGAGTGATTTGACCTGTTTCAGATTTTTCATACCGCTTCGCGTTATGTT 428
Db 348 CGACGGGCGACA---CGTTGCACTCTGTGCGGACTTCGCTACCGCTGCGCGGTATCGT 404
QY 429 GATATCTGAGCTGTGGAGTGCTTTCAGCGCAGATGGAAGATTTAAGCATGCTTGA 488
Db 405 GATCGCGAATCTCTGGCGGTGCGGCCCGCCGACCGCACCTGTTCCTGCTGCTGCGA 464
```

```
QY 489 TCTTCTGGTCAGTACACCG-----AAGGATAAAAGTGAAGAGTGA 530
Db 465 CCGGATGCTGCAGATCAGGTGCGCGGACGATGACGTTCGGCGACGACGCGCA 524
QY 531 AAAGGCTTTTGGAAAGACGAGATAGGTGTGAGGAGAACTGGCGCGGTTTTTGGCGG 590
Db 525 CGAGGACTACCAACGCTGCTCAAGAAACCCATGCGGCATGACGCGCTTACCTCCAGCA 584
QY 591 CATCATAGAGAAACCGGAAACAAACCGGACAGGATATTATTTCTATTTAGTGGAGC 650
Db 585 CCAAGTCAACGACCGCGCGCGCCCGCGGAAAGACCTGATCTCGGCACTCGTGGCGCG 644
QY 651 GGAAGAAACAGGCGAGAGCTGTCCGTTGAAGAGCTGATTTCCGTTTTTGCACGCTGTGCT 710
Db 645 CCGGTTGAGGCGCGAAGCACTCACCGACGAGAGATCGTGAATTCGGGCGCTGTGCT 704
QY 711 GGTGGCGGAAATGAAACCACTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAGA 770
Db 705 GATGGCGGCGACGCTCTCCACCTCCATGCTGTGCGCAACACCGCTCTGTGCTGAAGCA 764
QY 771 AACGCCAGCGGTTTACGAGGAACTGCGAGCAACCTGCGAGCCATCTGAACTGATGCTCAGGCACTGGA 830
Db 765 CCACCCCGCGGCGAGGCGCGCGCGCGCGGCTCCCTGATCCCGCCCTGATCGA 824
QY 831 GGAAGCTTTGCTTTTCAGAGCGCGCGCGCGGTTTTTGAAGCGCATTCGCCAAGCGGATAC 890
Db 825 AGAAGTACTGCGGCTGCGGCGCGCGATCACCGTTCATGGCGCGGCTCACCAAGGACAC 884
QY 891 GGAGATCGGGGGGCGACCTGATTAAGAAAGTGATATGTTTGGCGTTTTTGTGTCATCGGC 950
Db 885 CGTCTCGCGCGCACCCACCATCCCGCGGAGCGCATGTCGTGCGCTCCCTGCTGTCGCG 944
QY 951 AAATCGTGATGAAGCAAAAGTTTGACAGACCGCACATGTTTGATATCCGCGCCATCCCA 1010
Db 945 CAACACAGCAAGACAGGTTTTCACGACCCCGACACCTCGACCTCGCGCGGAAAGC-- 1002
QY 1011 TCCGATATTTGGTTTGGCCAGCGCATCCATTTTGGCTTGGGCGCGCGCTTGCCTGCT 1070
Db 1003 -CGCCAGATCGCTTTCGGCCAGCGCATCCACTACTGCTGGGCGCGCGCTCGCCGCT 1061
QY 1071 TGAAGCAATATCGC 1085
Db 1062 GGAGGCGCGCATCGC 1076
```

## RESULT 8

```
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Mutasynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6
```





Db 1068 TCTCGGCTACGGCTTCGGCGTGCACACTACTGCTGGCGCCTCGCTGCGCCGCGTGGAGGG 1127  
Qy 1077 AATATGCGGTTACGCTTTTGATTTCTCTCTTCTCATATGGAGT 1123  
Db 1128 GCGATGCCCATCAGCGCTGCTCGCGGCTTCCCGACCTCCAGT 1174

RESULT 10  
US-09-252-991A-1338/c  
; Sequence 1338, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1338  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1338

Query Match 9.2%; Score 109.8; DB 4; Length 1482;  
Best Local Similarity 47.6%; Pred. No. 8.4e-26;  
Matches 422; Conservative 0; Mismatches 432; Indels 33; Gaps 2;

Qy 240 CATTATATCATGACCGCGGAGCAATCAAAATCCGTTCACTGCTGGAACAAAGCCTT 299  
Db 1189 CATGCTCAACCTGACCGCGGAGCAACGCGGAGCAATCAAGATCAACGAGGAATGCT 1130

Qy 300 TACTCCGCGGCTGATGAAGCAATGGGACCGAGAAATCAAGAAATCAAGATGAATGAT 359  
Db 1129 CACCCCGCGCAGGTGGAGCGCTGCAACCGCATATAGAACGATCAACGAGGAATGCT 1070

Qy 360 TCAAAATTTACGGGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 419  
Db 1069 GAGCGGCTATGCGCGCGCGGAGCAACGCGGAGCAATCAAGATCAACGAGGAATGCT 1010

Qy 420 GGTATTGTGATATCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 479  
Db 1009 CATCGCGGTGATCTTTCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 950

Qy 480 ATGCTCTGATCTTCTGCTCAGTACACCGCAAGGATATAAAGTGAAGAGCTGAAAGAGCCTT 539  
Db 949 GTCTTGGAGCGCCAG-----GCGGAATGCTGTC 920

Qy 540 TTTGGAAGACGAGATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 599  
Db 919 GCGGAGGAGGCGCGGAGCGCTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 860

Qy 600 AGAAAGCGAAACAAACCGCAACGAGTATATTTCTATTATTGTAAGCGGAGGAAC 659  
Db 859 GCGCAGCGCGCGGAGCGCAACCGGAGCGCTGACAGCGGCGTGTGAGCGCGCGGAGCA 800

Qy 660 AGCGGAGAGCTGTCCGCTGGAAGAGCTGATTCGCTTTTGGACGCTGCTGCTGCTGCTGCTG 719  
Db 799 GAGCGGCAACTGAGCGAGCGGAGCTGCTCTCCATGCGCCACCTGCTGATGATGAGCGG 740

Qy 720 AATGAAACCACTACAAACCTGATTTCAATGCGATGTAGCATATATTAGAAACGCGCAGG 779  
Db 739 CTTTCGAGACCAACCATGATGATGCGCAACGCGCTGTGTCACTGCTGTGTGTGTGTGTGTGT 680

Qy 780 CTTTACGAGGAGTGTGCGGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 839  
Db 679 GCAACTGCGGCTGCTGCGGCGGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 620

Qy 840 GCGTTTCAGAGCGCGCGCCCGCTTTTTCAG---GCGATTCGCAAGCGGGATACGAGAT 896  
Db 619 CCGCCACGACAGCGCGCTGCGCGCTCGATGTTGCGCTTCACCGTGAAGAGCTGGAACCT 560

Qy 897 CGCGGGGCACTGATTAAGAAGCTGATATGTTTGGCGTTTGGCGATCGGCAATCG 956  
Db 559 GAGCGGGTCAACATTTCCGCGCGGAGATACATCTGCTCTCAACCTGACGCGCCACCA 500

Qy 957 TGATGAAGCAAGTTTGACAGACCGCACATGTTTGATATCCGCGCCATCCCAATCCCA 1015  
Db 499 CGAGCGCGAGCGCTTCGACGATCCGACCGCTCGACCTCACCGCAACACGATGGCCA 440

Qy 1017 TATTGCGTTTGGCAACGCGCATCCATTTTTCGCTGGGCGCGCGCTGCTGCGCTTGAASC 1076  
Db 439 TCTCGGCTAOCGCTTCGCGCGTGCACACTACTGCGTGGCGGCTCTGCTGGCGCGCTGGAGGG 380

Qy 1077 AATATCGGCTTACGCTTTTGATTTCTGCTTTTCTCATATGAGT 1123  
Db 379 GCGGATCGCATCCAGCGCGCTGCTCGCGGCTTCCCGACCTCCAGT 333

RESULT 11  
US-09-252-991A-1222  
; Sequence 1222, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1222  
; LENGTH: 1671  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1222

Query Match 9.2%; Score 109.8; DB 4; Length 1671;  
Best Local Similarity 47.6%; Pred. No. 9.1e-26;  
Matches 422; Conservative 0; Mismatches 432; Indels 33; Gaps 2;

Qy 240 CATCATTAACATGACCGCGCGGAGCAATCAAAATCCGTTCACTGCTGGAACAAAGCCTT 299  
Db 253 CATGCTCAACCTGACCGCGCGGAGCAATCAACCGCTGCGCTGCGTGGCGCGCGCTT 312

Qy 300 TACTCCGCGCTGATGAAGCAATGGGACCGAGAAATCAAGAAATCAAGATGAATGAT 359  
Db 313 CACCCCGCGCAGGTGGAGCGCTGCAACCGCATATAGAACGATCAACGAGGAATGCT 372

Qy 360 TCAAAATTTTCAGGGGCGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 419  
Db 373 GAGCGCATGCGCGCGCGCAACGAGCGCACTGATGCGCGACTTCGCGATCCGCGTAC 432

Qy 420 GGTATTGTGATATCTGAGCTGCTGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 479  
Db 433 CATCGGCTGATCTTCGAGCTGCTGGGCAATTCGCGAGCGCGAGCGCAACACGCGCGCA 492

Qy 480 ATGCTCTGATCTTCTGCTCAGTACACCGAAGGATAAAGTGAAGAGCTGAAAGAGCCTT 539  
Db 493 GTCTTGGAGCGCCAG-----GCGGAATGCTGTC 522

Qy 540 TTTGGAAGACGAGATGATGAGGAGAACTGCGCGCGTGTGCGGCAATATAGA 599  
Db 523 GCGGAGGAGCGCGCGCGCTGCGCGATGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 582

Qy 600 AGAAAGCGAAACAAACCGGAAACGAGGATATTATTCTTATTATTAGTGAAGCGGAGAAC 659

583 GGCACAGCGCCGCGCAGCAGCGCTCTACAGCGGGCTGTGTGACAGCGCGCGCGCA 642  
660 AGCGAGAGCTGTCGCGTGAAGAGCTGATTCGGTGGTTCAGCGCTGCTGCTGCGCGG 719  
643 GAGCGCGCACTGAGCGAGCGGGAACCTGCTCCATGCGCCACCTGCTGATGATGAGCGG 702  
720 AAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAGAAACGCCAGG 779  
703 CTTGAGACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 762  
780 CGTTTACAGAGAGCTGCGAGCCTGCTGATGATGATGATGATGATGATGATGATGATGAT 839  
763 GCAATGCGCTGCTGCGCGCGCAGCGGGAACCTGCTCCATGCGCCACCTGCTGATGATGAG 822  
840 GCGTTTACAGAGCGCGCGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 896  
823 CCGCACAGCAGCGCGCGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 882  
897 CGCGCGCGCAGCTGATTAAGAGAGTGTATGATGATGATGATGATGATGATGATGATGAT 956  
883 GACCGGGTCAACATTCGCGCGCGGGAATACATCTGCTTCAACCTGACCGCGCAACA 942  
957 TGATGAAGCAAGTTTACAGAGCGCAATGTTGATGATGATGATGATGATGATGATGATGAT 1016  
943 CAGCGCGAGCGCTTGCAGAGTCCGACCGCGCTGACCTCACCGCAACACCGATGCGCA 1002  
1017 TATTGCGTTGCGCACGCGCATTCATTTTGTGCTTGGCGCGCGCTGCTGCTGCTGCTGCT 1076  
1003 TCTGCGTACGCTTGGCGGTCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062  
1077 AAATATCGCGTTAAGCTTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1123  
1063 GCGGATGCCATCCAGCGCTGCTGCGCGCTTCCCGACCTCCAGT 1109

RESULT 12  
US-09-029-603-4  
; Sequence 4, Application US/09029603  
; Patent No. 6210935  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Engel, Natalie  
; APPLICANT: Bietenhader, Jurg  
; APPLICANT: Toupet, Christine  
; APPLICANT: Pospiech, Andreas  
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters  
; FILE REFERENCE: 4-20555/A/PCT  
; CURRENT FILING DATE: 1998-03-20  
; EARLIER APPLICATION NUMBER: PCT/EP96/03643  
; EARLIER FILING DATE: 1996-08-19  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 6085  
; TYPE: DNA  
; ORGANISM: Streptomyces longisporoflavus  
; FEATURE:  
; NAME/KEY: misc RNA  
; LOCATION: (378)..(1665)  
; OTHER INFORMATION: ORF  
; FEATURE:  
; NAME/KEY: misc RNA  
; LOCATION: (1747)..(2553)  
; OTHER INFORMATION: ORF  
; FEATURE:  
; NAME/KEY: misc RNA  
; LOCATION: (2593)..(4011)  
; OTHER INFORMATION: ORF  
; FEATURE:  
; NAME/KEY: misc RNA  
; LOCATION: (4013)..(4999)  
; OTHER INFORMATION: ORF

FEATURE:  
; NAME/KEY: misc RNA  
; LOCATION: (5071)..(6085)  
; OTHER INFORMATION: ORF  
US-09-029-603-4  
Query Match 8.7%; Score 103.6; DB 3; Length 6085;  
Best Local Similarity 51.0%; Pred. NO. 2.3e-23;  
Matches 244; Conservative 0; Mismatches 234; Indels 0; Gaps 0;  
QY 623 AGGATATTATTCTATTATTAGTGAAGCGGAGAACAGGCGAGAGCTGTCCGCTGAAG 682  
Db 1048 AGGATCTGCTACCTCTCTCTGTCGCGCGCGGACACCGGATCACCGCTCAGCGTGGAG 1107  
QY 683 AGCTGATTCGCTTTTGCACGCTGCTGCTGGTGGCGGAAATGAAACCACTACAAACCTGA 742  
Db 1108 GCATGCTCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1167  
QY 743 TTTCAAATGCGATGTACAGCATATTAGAAACGCCAGCGGCTTTTACGAGGAATCGCGAGCC 802  
Db 1168 TCGCCAGGCGGCTCTCTACCTGCGCGCGCCACCTGACGCTGCTGCTGCTGCTGCTGCTG 1227  
QY 803 ATCTGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862  
Db 1228 CACCGAGTGTGACACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1287  
QY 863 TTTTGAGGCGCATTCGCAAGCGGATACGAGATCGGGGGGCGCTGATTAAGAGAGGTG 922  
Db 1288 CGTACGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1347  
QY 923 ATATGCTTTTGGCGTTTGTGCGCATCGGCAATCGTGTGATGATGATGATGATGATGATG 982  
Db 1348 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1407  
QY 983 ACATGTTGATATCCGCGCCATCCCAATCCGATATTCGCTTTGCGCCACGCGATCCATT 1042  
Db 1408 AGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467  
QY 1043 TTTGCTTTGGCGCGCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 1100  
Db 1468 ACTGCTGCGCGGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1525

RESULT 13  
US-09-266-965-24  
; Sequence 24, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24  
; LENGTH: 1233  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-24  
Query Match 7.5%; Score 88.8; DB 4; Length 1233;  
Best Local Similarity 51.1%; Pred. NO. 6.8e-19;

Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCGGTGAAGAGCTGATTCGGTTTGCA 700  
DB 665 TCCTGACCGCGGAGAGAGAGAGCTCAGAGGGGAGCTGGTGGCTGGCGG 724  
QY 701 CGCTGCTGCTGGTGGCGGAGAAATGAACCACTACAAACCTGATTTCAAAATCGGATGACA 760  
DB 725 AGGCCATCTCTCATCGCCGGCTACGAGACCTCGGCGAGCCAGATCCCCAACTTCCTCTACG 784  
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTCGCAGCCATCCTGAACTGATGCTTC 820  
DB 785 TCCTCTTCGCGCCACCGCAGCTGCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 844  
QY 821 AGGCAGTGGAGAGACCTTGGCTTTTCAGAGCGCGGCGCCCGCTT-----TTGAGGCGCA 874  
DB 845 ACGCGCTCGAGGAACCTGCTGGCTTCGTGCCATCGGCACCGCTGACCGCTTTCGCCGTA 904  
QY 875 TTGCAAGCGGATACGAGATCGGGGGACCTGATTAAGAGAGGATGATGTTTGG 934  
DB 905 CGGCCACCGAGGACCTCGAGCTCGGGGGAGTCTCTGGTCAGGGCGGGGAGACGCTCGTGC 964  
QY 935 CGTTTGGCATCGCAATCGTATGAGCAAAATTTGACAGACCGCACATGTTTGATA 994  
DB 965 CGTCGATGGGCGCGCAACCGCACCCGAGCTGTTACGAGACCGCGAGCTGGACC 1024  
QY 995 TCCGCGCCCATCCCAATCGCATATGCTGTGGCCACGCGATCAATTTTSCCTTGGGG 1054  
DB 1025 TCGCGCGCGCGCAATCGCACTCGGCTTCGGCGGGGACCGCACCTGCTGGCGG 1084  
QY 1055 CCGCGCTGCGCGCTTTGAACCAATATCGGCTTAACGCTTTTGAT 1100  
DB 1085 CCAACTGGCGCGGCTGGAGCTCCAGATCAGCTCAGCACGCTGTT 1130

RESULT 14

US-09-266-965-74

; Sequence 74, Application US/09266965

; Patent No. 6495348

; GENERAL INFORMATION:

; APPLICANT: Sherman, D

; APPLICANT: Mao, Y

; APPLICANT: He, M

; APPLICANT: Varoglu, M

; APPLICANT: Sheldon, P

; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster

; FILE REFERENCE: 600.456US1

; CURRENT APPLICATION NUMBER: US/09/266,965

; EARLIER FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 08/624,447

; EARLIER FILING DATE: 1996-08-19

; EARLIER APPLICATION NUMBER: PCT/US94/11279

; EARLIER FILING DATE: 1994-10-06

; EARLIER APPLICATION NUMBER: US 08/133,963

; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 74

; LENGTH: 12249

; TYPE: DNA

; ORGANISM: Streptomyces lavendulae

US-09-266-965-74

Query Match 7.5%; Score 88.8; DB 4; Length 12249;  
Best Local Similarity 51.1%; Pred. No. 2.9e-18;  
Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCGGTGAAGAGCTGATTCGGTTTGCA 700  
DB 9239 TCGTGACCGCGCGGAGAGAGCAAGCTCAGCGAGGCGAGCTGGTGGCTGGCGG 9298  
QY 701 CGCTGCTGCTGGTGGCGGAGAAATGAACCACTACAACTGATTTCAAAATGCGATGACA 760

DB 9299 AGGCATCTCTCATCGCGGCTACGAGACCTCGGCGAGCGAGTCCCAACTTCCTCTACG 9358  
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTGCGAGGCATCCTGAACCTGATGCTTC 820  
DB 9359 TCCTCTTCGCGCCACCGCAGCTGCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 9418  
QY 821 AGGCAGTGGAGAGACCTTGGCTTTCAGAGCGCGCGCCCGCTT-----TTGAGGCGCA 874  
DB 9419 ACGCGCTCGAGGAACCTGCTGGCTTCGTCGCCATCGGCACCGTGGACCGCTTTCGCCGTA 9478  
QY 875 TTGCAAGCGGAGATACGAGATCGGGGGCACCTGATTAAGAGAGCTGATGTTTGG 934  
DB 9479 CGGCCACCGAGAGCTCGAGCTCGGGGAGTCTCTGGTCAGGGCGGGGAGACGCTCGTGC 9538  
QY 935 CGTTTGGCATCGCAATCGTATGAGCAAAATTTGACAGACCGCACATGTTTGATA 994  
DB 9539 CGTCGATGGGCGCGCAACCGCACCCGAGCTGTTACGAGACCGCGAGCTGGACC 9598  
QY 995 TCCGCGCCCATCCCAATCGCATATGCTGTGGCCACGCGATCAATTTTSCCTTGGGG 1054  
DB 9599 TCGCGCGCGCGCAATCCGCACTCGGCTTGGCGCGGAGCCGACCACTGCTGGCGG 9658  
QY 1055 CCGCGCTGCGCGCTTTCGAAGCAAAATATCGGCTTAACGCTTTTGAT 1100  
DB 9659 CCAACTGGCGCGGCTGGAGCTCCAGTCACTCAGCTCAGCACGCTGTT 9704

RESULT 15

US-09-266-965-96

; Sequence 96, Application US/09266965

; Patent No. 6495348

; GENERAL INFORMATION:

; APPLICANT: Sherman, D

; APPLICANT: Mao, Y

; APPLICANT: Varoglu, M

; APPLICANT: He, M

; APPLICANT: Sheldon, P

; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster

; FILE REFERENCE: 600.456US1

; CURRENT APPLICATION NUMBER: US/09/266,965

; EARLIER FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 08/624,447

; EARLIER FILING DATE: 1996-08-19

; EARLIER APPLICATION NUMBER: PCT/US94/11279

; EARLIER FILING DATE: 1994-10-06

; EARLIER APPLICATION NUMBER: US 08/133,963

; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 96

; LENGTH: 18331

; TYPE: DNA

; ORGANISM: Streptomyces lavendulae

US-09-266-965-96

Query Match 7.5%; Score 88.8; DB 4; Length 18331;  
Best Local Similarity 51.1%; Pred. No. 3.8e-18;  
Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCGGTGAAGAGCTGATTCGGTTTGCA 700  
DB 9239 TCGTGACCGCGCGGAGAGCAAGCTCAGCGAGGAGCAAGCTCAGCGAGGAGCTGGTGGCTGGCGG 9298  
QY 701 CGCTGCTGCTGGTGGCGGAGAAATGAACCACTACAACTGATTTCAAAATGCGATGACA 760  
DB 9299 AGGCATCTCTCATCGCGCTACGAGACCTCGGCGAGCGAGTCCCCAACTTCCTCTACG 9358  
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTCGGAGCCATCCTGAACCTGATGCTTC 820  
DB 9359 TCCTCTTCGCGCCACCGCAGCTGCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 9418  
QY 821 AGGCAGTGGAGAGACCTTGGCTTTCAGAGCGCGCGCCCGCTT-----TTGAGGCGCA 874

Db 9419 ACGCCGTCGAGAACTGCTGCGCTTCGTGCCCATCGGCACCGTGGACGGCTTTCGCCGTA 9478  
Qy 875 TTGCCAAGCGGGATACGAGATCGGGGGCACCTGATTAAAGAGGTGATATGGTTTGG 934  
Db 9479 CGGCCACCGAGGAGTCTGAGCTCGGGGAGTCTTGTCAGGGCCGGGAGACGGTCGTGC 9538  
Qy 935 CGTTTGTGGCATCGSCAAATCGTGATGAAGCAAAAGTTTGAAGACCCGCCACATGTTTGATA 994  
Db 9539 GTCGATGGGCGCGCCCAACCGGACCCCGAGCTGTTACGAGACCCCGACGAGCTGGACC 9598  
Qy 995 TCCGCGGCCATCCCAATCCGCAATATGCGTTTGGCCACGGCATCCATTTTGGCTTGGGG 1054  
Db 9599 TCGGCGGGCGCGCAATCCGACCTTGGGCTTGGGCGCGGACCGCACCACTGCCCTGGGCG 9658  
Qy 1055 CCGCGTTGCCCGCTCTTGAAGCAATATCGGTTAAACGTCITTGAT 1100  
Db 9659 CCCAACTGGCCCGGGTGGAGCTCCAGATCACGCTCACGACGCTGTT 9704

Search completed: May 28, 2004, 22:10:36  
Job time : 108 secs

> O <  
O | O IntelliGenetics  
> O <

GENALIGN - Multiple Sequence Alignment Program  
Release 5.4

Wed 2 Jun 104 12:59:50-PDT

Solution Parameters:

Nucleic Alphabet = Identity  
Output line length = 80  
Compress = Off  
Histogram = Off  
Randomization = Off  
AMINO-Res-length = 2  
Deletion-weight = 5.00  
Length-factor = 0  
Matching-weight = 1.00  
NUCLEIC-Res-length = 4  
Spread-factor = 50

Clustered order of selected sequences:

1. US-09-869-334B-2 (1-1191)  
40. US-09-869-334B-41 (1-1204)  
41. US-09-869-334B-43 (1-1221)  
42. US-09-869-334B-44 (1-1221)

Region Alignment: (listed in Clustered order)

```

US-09-869- 1      ATGAATGTTAAACCGCGCGCAAGCCTTGCGAGCGAG
US-09-869- 1      GTCGAC  aATGAATGTTAAACCGCGCGCAAGCCTTGCGAGCGAG
US-09-869- 1      CTCGAGTCGAGGAGTCGACTAATATGAACGTTCTGAACCGCGCGTCAAGCCTTGCGAGCGAG
US-09-869- 1      CTCGAGTCGAGGAGTCGACTAATATGAACGTTCTGAACCGCGCGTCAAGCCTTGCGAGCGAG
consensus      ctcgagtcgagaggtcgactaatatgaa-gt--T-AACCGCGG-CAAGCCTTGCGAGCGAG

US-09-869- 38      CGCTGCTCAATGGGAAAAACAACAGGATCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 45      CGCTGCTCAATGGGAAAAACAACAGGATCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 62      CGCTGCTCAATGGGAAAAACAACAGGATCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 62      CGCTGCTCAATGGGAAAAACAACAGGATCGTATCATCCGTTTCATGGTATGAATCGAT
consensus      CGCTGCTCAATGGGAAAAACAACAGGATCGTATCATCCGTTTCATGGTATGAATCGAT

US-09-869- 99      GAGAAAGGATCGCCTGTTTCCCTTGTATGAAGAAAAACAAGTGTGAGCGGTTTTTCTTTAT
US-09-869- 106     GAGAAAGGATCGCCTGTTTCCCTTGTATGAAGAAAAACAAGTGTGAGCGGTTTTTCTTTAT
US-09-869- 123     GAGAAAGGATCGCCTGTTTCCCTTGTATGAAGAAAAACAAGTGTGAGCGGTTTTTCTTTAT
US-09-869- 123     GAGAAAGGATCGCCTGTTTCCCTTGTATGAAGAAAAACAAGTGTGAGCGGTTTTTCTTTAT
consensus      GAGAAAGGATCGCCTGTTTCCCTTGTATGAAGAAAAACAAGTGTGAGCGGTTTTTCTTTAT

US-09-869- 160     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCAGA
US-09-869- 167     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCAGA
US-09-869- 184     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCAGA

```

```

US-09-869- 184  GATGATGTCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCAGA
consensus      GATGATGTCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCAGA

US-09-869- 221  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 228  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 245  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 245  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
consensus      CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC

US-09-869- 282  AGTCGTGAAACAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGAA
US-09-869- 289  AGTCGTGAAACAAGCCTTTACTCCGCGCGGATGAAGCAATGGGAACCGGAATTCAGAA
US-09-869- 306  AGTCGTGAAACAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGAA
US-09-869- 306  AGTCGTGAAACAAGCCTTTACTCCGCGCGGATGAAGCAATGGGAACCGGAATTCAGAA
consensus      AGTCGTGAAACAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGAA

US-09-869- 343  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGATT
US-09-869- 350  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGATT
US-09-869- 367  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGATT
US-09-869- 367  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGATT
consensus      ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGATT

US-09-869- 404  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTSCCTTTCAGGCGAGAT
US-09-869- 411  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTSCCTTTCAGGCGAGAT
US-09-869- 428  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTSCCTTTCAGGCGAGAT
US-09-869- 428  ATTTCATACCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTSCCTTTCAGGCGAGAT
consensus      TTTTCATACCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTSCCTTTCAGGCGAGAT

US-09-869- 465  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTGATACACCGAAGGATAAAAGTGAAGAA
US-09-869- 472  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTGATACACCGAAGGATAAAAGTGAAGAA
US-09-869- 489  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTGATACACCGAAGGATAAAAGTGAAGAA
US-09-869- 489  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTGATACACCGAAGGATAAAAGTGAAGAA
consensus      GGAACAGTTTAAAGCATGCTCTGATCTTCTGTGATACACCGAAGGATAAAAGTGAAGAA

US-09-869- 526  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGG
US-09-869- 533  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGG
US-09-869- 550  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGG
US-09-869- 550  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGG
consensus      GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGG

US-09-869- 587  CCGGCATCATAGAAAGCGAAACAAACCGGAACAGGATATTATTTCTATTTAGTGA

```

```
US-09-869- 594 CCGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
US-09-869- 611 CCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
US-09-869- 611 CCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
consensus
|
|
|
US-09-869- 648 AGCGGAAGAAACAGCGGAGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCACGCTGCTG
|
|
|
US-09-869- 655 AGCGGAAGAAACAGCGGAGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCACGCTGCTG
|
|
|
US-09-869- 672 AGCGGAAGAAACAGCGGAGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCACGCTGCTG
|
|
|
US-09-869- 672 AGCGGAAGAAACAGCGGAGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCACGCTGCTG
|
|
|
consensus
|
|
|
US-09-869- 709 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 716 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 733 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 733 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
consensus
|
|
|
US-09-869- 770 AAACGCCAGGCGTTTACAGAGAACTGCGAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 777 AAACGCCAGGCGTTTACAGAGAACTGCGAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 794 AAACGCCAGGCGTTTACAGAGAACTGCGAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 794 AAACGCCAGGCGTTTACAGAGAACTGCGAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
consensus
|
|
|
US-09-869- 831 GGAAGCCTTCGGTTTACAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 838 GGAAGCCTTCGGTTTACAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 855 GGAAGCCTTCGGTTTACAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 855 GGAAGCCTTCGGTTTACAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
consensus
|
|
|
US-09-869- 892 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGGTTTTTGGCGTTTGTGGCATCGGCAG
|
|
|
US-09-869- 899 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGGTTTTTGGCGTTTGTGGCATCGGCAG
|
|
|
US-09-869- 916 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGGTTTTTGGCGTTTGTGGCATCGGCAG
|
|
|
US-09-869- 916 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGGTTTTTGGCGTTTGTGGCATCGGCAG
|
|
|
consensus
|
|
|
US-09-869- 953 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCATCCCAATCC
|
|
|
US-09-869- 960 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCATCCCAATCC
|
|
|
US-09-869- 977 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCATCCCAATCC
|
|
|
US-09-869- 977 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCATCCCAATCC
|
|
|
consensus
|
|
|
```

```
US-09-869- 1014 GCATATTGCGTTTGCCACCGCATCCATTTTTCCTTGGGGCCCGCTTGCCCGTCTTGAA
|
|
|
US-09-869- 1021 GCATATTGCGTTTGCCACCGCATCCATTTTTCCTTGGGGCCCGCTTGCCCGTCTTGAA
|
|
|
US-09-869- 1038 GCATATTGCGTTTGCCACCGCATCCATTTTTCCTTGGGGCCCGCTTGCCCGTCTTGAA
|
|
|
US-09-869- 1038 GCATATTGCGTTTGCCACCGCATCCATTTTTCCTTGGGGCCCGCTTGCCCGTCTTGAA
|
|
|
consensus
|
|
|
US-09-869- 1075 GCAAAATATGCGGTAAACGTCTTTGATTTCTCTTTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1082 GCAAAATATGCGGTAAACGTCTTTGATTTCTCTTTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1099 GCAAAATATGCGGTAAACGTCTTTGATTTCTCTTTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1099 GCAAAATATGCGGTAAACGTCTTTGATTTCTCTTTCATATGAGAGTGGGTCAAGTATCA
|
|
|
consensus
|
|
|
US-09-869- 1136 CTCGANTTGAACACTGTGTATATACGGATTAAAGAGCTTCGGTGTGAAATGTAA
|
|
|
US-09-869- 1143 CTCGANTTGAACACTGTGTATATACGGATTAAAGAGCTTCGGTGTGAAATGTAAAGATC
|
|
|
US-09-869- 1160 CTCGANTTGAACACTGTGTATATACGGATTAAAGAGCTTCGGTGTGAAATGTAAAGATC
|
|
|
US-09-869- 1160 CTCGANTTGAACACTGTGTATATACGGATTAAAGAGCTTCGGTGTGAAATGTAAAGATC
|
|
|
consensus
|
|
|
US-09-869- 1192
|
US-09-869- 1204 C
|
US-09-869- 1221 C
|
US-09-869- 1221 C
|
consensus
|
|
|
Alignment score = 3366.00
Scoring matrix:
```

	1	40	41	42
1				
40	1123	1035	1027	
41		1111	1107	1213
42				

> O <  
O|||o IntelliGenetics  
> O <

GENALIGN - Multiple Sequence Alignment Program  
Release 5.4

Wed 2 Jun 104 12:57:39-PDT

Solution Parameters:

Amino Alphabet = Identity  
Output line length = 80  
Compress = Off  
Histogram = Off  
Randomization = Off  
  
AMINO-Res-length = 2  
Deletion-weight = 5.00  
Length-factor = 0  
Matching-weight = 1.00  
NUCLEIC-Res-length = 4  
Spread-factor = 50

Clustered order of selected sequences:

- 1. US-09-869-334B-1 (1-396)
- 2. US-09-869-334B-42 (1-396)
- 3. US-09-869-334B-45 (1-396)

Region Alignment: (listed in Clustered order)

```
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPFPWYESMRKDAPVSFDEENQVMSVFLYDDVKKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPFPWYESMRKDAPVSFDEENQVMSVFLYDDVKKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPFPWYESMRKDAPVSFDEENQVMSVFLYDDVKKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPFPWYESMRKDAPVSFDEENQVMSVFLYDDVKKVVG
consensus MNVLNRRQALQALLNGKKNQDAYHPFPWYESMRKDAPVSFDEENQVMSVFLYDDVKKVVG

US-09-869- 62 DKELFSSCMPQQTSSIGNSIINMDPKHTKIRSVVNKAFTPRAMKQWEPRIQIITDELIQ
US-09-869- 62 DKELFSSCMPQQTSSIGNSIINMDPKHTKIRSVVNKAFTPRAMKQWEPRIQIITDELIQ
US-09-869- 62 DKELFSSCMPQQTSSIGNSIINMDPKHTKIRSVVNKAFTPRAMKQWEPRIQIITDELIQ
consensus DKELFSSCMPQQTSSIGNSIINMDPKHTKIRSVVNKAFTPRAMKQWEPRIQIITDELIQ

US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQPKAWSDDLIVSTPKDKSBEAEKAFLEE
US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQPKAWSDDLIVSTPKDKSBEAEKAFLEE
US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQPKAWSDDLIVSTPKDKSBEAEKAFLEE
consensus FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQPKAWSDDLIVSTPKDKSBEAEKAFLEE

US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGSEELIPfCTLLLVAGNETT
US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGSEELIPfCTLLLVAGNETT
US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGSEELIPfCTLLLVAGNETT
consensus RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGSEELIPfCTLLLVAGNETT

US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIIGHLLIK
US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIIGHLLIK
```

```
US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIIGHLLIK
consensus TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIIGHLLIK

US-09-869- 306 EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
US-09-869- 306 EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
US-09-869- 306 EGDVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
consensus EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL

US-09-869- 367 ISAFPHMECVSITPIENSVIYGLKSPRVKM
US-09-869- 367 ISAFPHMECVSITPIENSVIYGLKSPRVKM
US-09-869- 367 ISAFPHMECVSITPIENSVIYGLKSPRVKM
consensus ISAFPHMECVSITPIENSVIYGLKSPRVKM
```

Alignment score = 785.00

Scoring matrix:

	1	2	3
1		394	389
2			391
3			

-----



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:04:11 ; Search time 71 seconds  
(without alignments)  
536.505 Million cell updates/sec

Title: US-09-869-334B-1  
Perfect score: 2052  
Sequence: 1 MNVLRQALORALLGNKK.....SITPIENSVYGLKSFVRVM 396

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:.\*  
1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	100.0	396	1 B69851	cytochrome P450 Yj
2	738	36.0	410	2 S39924	cytochrome P450meg
3	711.5	34.7	410	1 O4BS6M	cytochrome P450 10
4	674	32.8	417	1 S49051	cytochrome P450 ty
5	658	32.1	410	1 S49611	cytochrome P450 cy
6	646.5	31.5	405	1 B42606	cytochrome P450 CV
7	626	30.5	376	1 G69679	polyketide hydroxy
8	600	29.2	395	1 G69594	cytochrome P450 bi
9	581	28.3	397	1 S51594	cytochrome P450 my
10	581	28.3	434	2 C87170	probable cytochrom
11	571	27.8	418	2 G83229	cytochrome P450 PA
12	565.5	27.6	412	1 B40634	erythromycin monoo
13	564	27.5	404	2 T30231	cytochrome P450 -
14	548.5	26.7	368	1 S71328	cytochrome P450 CY
15	540	26.3	406	1 S18531	cytochrome P450 er
16	538	26.2	310	2 T44767	cytochrome P450 li
17	536.5	26.1	411	2 T36526	probable cytochrom
18	532.5	26.0	411	1 J25859	polyketide synthas
19	524	25.5	415	2 A27331	cytochrome P450 hy
20	524	25.5	415	2 B97512	probable cytochrom
21	520	25.3	428	1 A70729	cytochrome P450 Rv
22	516	25.1	402	1 A70707	cytochrome P450 Rv
23	514.5	25.1	398	1 H70807	cytochrome P450 Rv
24	499	24.3	433	1 B70677	cytochrome P450 Rv
25	497.5	24.2	444	2 F87366	cytochrome P450 fa
26	489.5	23.9	402	2 H82813	cytochrome P450-li
27	488	23.8	414	1 E70708	cytochrome P450 Rv
28	482	23.5	406	1 A48495	linalool 8-monooxy
29	481	23.4	411	1 J42403	cytochrome P450 -

ALIGNMENTS

RESULT 1

B69851  
Cytochrome P450 yj1B - Bacillus subtilis  
N:Contains: oxidoreductase (BC 1.-.-.-)  
C:Species: Bacillus subtilis  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #ext\_change 16-Jun-2000  
C:Accession: B69851  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.I.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapides, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.  
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: B69851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-396 <KJN>  
A:Cross-references: GB:299110; GB:AL009126; NID:526333472; PIDN:CAB13078.1; PID:92633575  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yj1B  
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
F:236-371/Domain: cytochrome P450 homology <CYP>

Query Match 100.0%; Score 2052; DB 1; Length 396;  
Best Local Similarity 100.0%; Pred. No. 1.6e-134;  
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNVLRQALORALLGNKKQDAYHPFPWYEMRKDAPVSFDEENQVNSVFLYDDVKVY	60
DB	1	MNVLRQALORALLGNKKQDAYHPFPWYEMRKDAPVSFDEENQVNSVFLYDDVKVY	60
QY	61	GDKELFSSCMPQOTSIGNSIINMDPPKTKRSVVKAFTRVVKMQRPRIOETDELI	120
DB	61	GDKELFSSCMPQOTSIGNSIINMDPPKTKRSVVKAFTRVVKMQRPRIOETDELI	120
QY	121	QKQGRSEFDLVHDFSYPLPVTIVISLGVPSAHMQFKANSDDLVSFPKSEAEKAF	180
DB	121	QKQGRSEFDLVHDFSYPLPVTIVISLGVPSAHMQFKANSDDLVSFPKSEAEKAF	180
QY	181	LEERDKCEELAAFPAGIIEBKRNKPEQDIISILVEAETGKLSGELLIPFTLLVAG	240

```
Db 181 LERDKOBEELAAFPNGIIIEERNKPEQDIISLVEAETGKLSGSELIIPFCTLLVAG 240
Qy 241 NETTNLSNAMYSLTPTGVYELRSHPELMPQAVEEALRPAPVLRRAKRTTEIG 300
Db 241 NETTNLSNAMYSLTPTGVYELRSHPELMPQAVEEALRPAPVLRRAKRTTEIG 300
Qy 301 GHLIEGDMVLAFVANSRDEAKFDRPHMFDIRRHNPFIAPGHGIFHFCGLGAPLARLEAN 360
Db 301 GHLIEGDMVLAFVANSRDEAKFDRPHMFDIRRHNPFIAPGHGIFHFCGLGAPLARLEAN 360
Qy 361 IALTSLISAPFPMCEVSIPTIENSIVYGLKSRVXM 396
Db 361 IALTSLISAPFPMCEVSIPTIENSIVYGLKSRVXM 396

RESULT 2
cytochrome P450meg - Bacillus megaterium (ATCC 13368)
C:Species: Bacillus megaterium
A:Variety: ATCC 13368
C:Date: 07-Oct-1994 #sequence_revision 23-Feb-1996 #text_change 28-Jul-2000
R:Auschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Siewert, G.
Mol. Gen. Genet. 241, 170-176, 1993
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid
A:Reference number: S39924; MUID:94049677; PMID:8232201
A:Accession: S39924
A:Molecule type: DNA
A:Residues: 1-410 <RAU>
A:Cross-references: EMBL:221972; NID:G288298; PIDN:CAA79985.1; PID:G288300
A:Experimental source: ATCC 13368
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:240-377/Domain: cytochrome P450 homology <P45>
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 36.0%; Score 736; DB 2; Length 410;
Best Local Similarity 41.1%; Pred. No. 1.3e-43;
Matches 157; Conservative 74; Mismatches 133; Indels 18; Gaps 5;

Qy 18 KKKQDAYHFPWYSGMRKQAPVDFEENQVMSVFLYDDYKVVGVGDKELFSSCMPQQTSSI 77
Db 14 KRTBEFGPYAKCKRMLENDPVSYHEGDTWNVFYEDVKVLSYKHFSSVRKRTTISV 73
Qy 78 GN-----SIINMDPPKHTKIRSVVNVKAFTRPVKQWEPRIQETIDELIQEQGR 126
Db 74 GDSSEGSVPEKIQTESDPPHRRKRSLLAAFTPSLQNWEPRIQETIDELIQEQGR 133
Qy 127 SEFDLVHDFSYPLPVIIVISLGLVPSAHMEQKANSDDLVS-TPKDKSEAEKAFLEERD 185
Db 134 TEIDIVASLASPLPIVWADLGVPSKORLLPKKWDTLFLPDRKQEQEVDKL-----KQ 189
Qy 186 KCEELAAFFAGIIEKRNKPEODIISILVEAETGKLSGSELIIPFCTLLVAGNETTT 245
Db 190 VAAKSYQYQYLPVIVQKRLNPADDIISLKSSEVDGEMFTDDEVATMLLILGAGVETTS 249
Qy 246 NLISNAMYSL-ETPGVYELRSHPELMPQAVEEALRPAPVLRRAKRTTEIGGHLI 304
Db 250 HLGANSFSLLYDDKEVYQELHNLDLVQAVEEMLRFRNLKLDRTVKEDNDLLGVGL 309
Qy 305 KEGDMVLAFVANSRDEAKFDRPHMFDIRRHNP-PIAFGHGIFHFCGLGAPLARLEANIAL 363
Db 310 KEGDSVVVWMSAANDMEEMFDPFTLINHRPNKHLTFTGNGPHFCLGAPLARLEAKIAL 369
Qy 364 TSLISAPFPMCEVSIPTIENSIV 385
Db 370 TAPLKKFKHIEAVSPSQLEENL 391

RESULT 3
Q4856M
cytochrome P450 106 - Bacillus megaterium
```

```
N:Alternate names: cytochrome P450BM-1
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus megaterium
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C:Accession: S07764; S17973
R:He, J.S.; Ruettinger, R.T.; Liu, H.M.; Fulco, A.J.
Biochim. Biophys. Acta 1009, 301-303, 1989
A:Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P-4
A:Reference number: S07764; MUID:90089408; PMID:2597681
A:Accession: S07764
A:Molecule type: DNA
A:Residues: 1-410 <HEJ1>
A:Cross-references: EMBL:X16610; NID:G39626; PIDN:CAA34612.1; PID:G39627
A:Accession: S17973
A:Molecule type: protein
A:Residues: 1-25 <HEJ2>
C:Genetics:
A:Gene: CYP106
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
F:1-410/Product: cytochrome P450 106 #status experimental <MAT>
F:241-378/Domain: cytochrome P450 homology <CYP>
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 34.7%; Score 711.5; DB 1; Length 410;
Best Local Similarity 39.7%; Pred. No. 9.1e-42;
Matches 151; Conservative 74; Mismatches 132; Indels 23; Gaps 6;

Qy 18 KKKQDAYHFPWYSGMRKQAPVDFEENQVMSVFLYDDYKVVGVGDKELFSSCMPQQTSSI 77
Db 15 QSRAEFFPIQWYKEMLNNSPVYFHEETNTWNVFYEHVQVLSNYDFFSSDQRTITFV 74
Qy 78 GN-----SIINMDPPKHTKIRSVVNVKAFTRPVKQWEPRIQETIDELIQEQGR 126
Db 75 GDSKKKSTSPITNLNLPDHRKARSLLAAFTPSLQNWEPRIKQIAADLVEALQKN 134
Qy 127 SEFDLVHDFSYPLPVIIVISLGLVPSAHMEQKANSDDLVS-TPKDKSEAEKAFLEERD 185
Db 135 STINIVDDLSPPSLVADLFGVPVKDRYQFKKWDILFPQYDQERLEIE----CEKQ 190
Qy 186 KCEELAAFFAGIIEKRNKPEODIISILVEAETGKLSGSELIIPFCTLLVAGNETTT 245
Db 191 RAGAEYQYLPVIVIEKRNLSODIISDLIQAEVDGETTDEIVHATMLLILGAGVETTS 250
Qy 246 NLISNAMYSL-ETPGVYELRSHPELMPQAVEEALRPAPVLRRAKRTTEIGGHLI 304
Db 251 HALANFYSLYDKSLYSELNRELAKPAVEEMLRYRHFHISRDRTVKQDNELLGVKL 310
Qy 305 KEGDMVLAFVANSRDEAKFDRPHMFDIRRHNP-PIAFGHGIFHFCGLGAPLARLEANIAL 363
Db 311 KKGDDVVVWMSACNDMETFENPFSDIHRPTNKKHLTFTGNGPHFCLGAPLARLEMKIL 370
Qy 364 TSLISAPFPMCEVSIPTIEN 383
Db 371 EAPLEAFSHIE-----PFED 385

RESULT 4
S49051
cytochrome P450 tyli (EC 1.1.1.-) [validated] - Streptomyces fradiae (strain T59235)
C:Species: Streptomyces fradiae
A:Variety: strain T59235
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S49051
R:Merison-Davies, L.A.; Cundliffe, E.
Mol. Microbiol. 13, 349-355, 1994
A:Title: Analysis of five tylosin biosynthetic genes from the tyliBA region of the Strept
A:Reference number: S49051; MUID:95075319; PMID:7984112
A:Accession: S49051
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-417 <MER>
A:Cross-references: EMBL:U08223; NID:G6849140; PIDN:AAA21341.1; PID:G473597
```



D6		63	RITEDKVNVMQEQIKMLNVSEDDIFVDSDHMLAKTDPDHTLRSLVHQAPPTPIENLRC	122
Q7		111	RIQBITDELIOKFQGRSEFDLVHDPSYPLPVIVISSELLGVPSAHMEQFKAWSDLLYSTPK	170
D6		123	SIEGTAEQLLDXEMKENKADIMKSPASPLPFTVISGELMGIPKEDRSQFOITNAMYDTISE	182
Q7		171	DKSEABKAFLEERDKCEEELAAFPAGIIEBKRNKEPDQIIISILVEAETGEKLSGSELI	230
D6		183	GNRELTNQALREFKD-----YIAKLHDDRIRKPKDDLISKLVHAENGSKLSKELY	234
Q7		231	PFCTLLLVAGNETTNLNISNAMYSILETGCVYEELRSHPELMQAVEALRFRAAPAVL-	289
D6		235	SMLFLVLVAGLETTVNLGLSGTLLAQHKKECEKUKQQPEMIATAVEEILRYTSPVVOMA	294
Q7		290	RRIAROTEIGCHLIKEGDMVLAFVASNRDEAKFDRPHMEDIRRHHPNHPIAFGHGIHFC	349
D6		295	NRWAIETYKGHSIKRGMDMFIGIGSARNDFNFENPEILNINSRPNRHISFGFGIHFC	354
Q7		350	LGAPLARLEANIA 362	
D6		355	LGAPLARLEGHIA 367	

cytochrome P450 bioI - *Bacillus subtilis*  
N:Contans: oxidoreductase [EC 1.-.-.-]  
C:Species: *Bacillus subtilis*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: G669594  
E:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, C.A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Feltz, B.; Gots, J.; Han, J.; Heilmann, S.; Hendrickson, E.; Hirsch, J.; Klenke, U.; Klotzel, J.; Kuhn, M.; Labat, A.; Laguerre, A.; Lecomte, L.; Leising, G.; Lin, J.; Little, J.; Liu, X.; Liu, Y.M.; Ogawa, K.; Ogatawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Poh, T.M.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaite, Y.; Sato, T.; Schaefer, M.; Schlegel, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Akeuchi, M.; Tanaka, K.; Tanaka, R.; Terpstra, P.; Tognoni, A.; Tosato, V.; T. Winters, P.; Wipat, A.; Yasumoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus*  
A:Reference number: A69580. MUPP-98044033. DWD-938477

AtcGene: b101  
 C-Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
 C-Keywords: oxidoreductase  
 F-232-367/Domain: cytochrome P450 homology <CYP>  
 F-232-367/Domain: cytochrome P450 homology <CYP>

QY	25	HPFPWYSMRKDAV----	SFDEENQVMSVFLYDDVKKVWQKEL-----	SSSCMPQI	74
Db	15	NPYSFYDTRAVHPZYKGSF-	LKYPGWVTVGYEETAALIKOARFVRTPLPSS	TKYQL	73
QY	75	SSIGNS-IINMDPKHTKIRSVNKAFTPRVW	KOWEPRIQETIDELIKFQGRSEFDLVH	133	
Db	74	SHVQNMFLQNGDHRRLPYLAGAFTPR	TESYQPIYIETVHLLDVQCKKMEVIS	133	
QY	134	DFSPLPVIVISLGLVPSAHMQKAWSDLLVST-	PKDKSEAEKAFLEERDKCEEELA	192	
Db	134	DFAPPLASFLVIANI:::GPEEDREOLKSWA	SLIQIDFTRSKA---LTEGINAVQAM	169	

[illegible]

RESULT 9  
S51594  
cytochrome P450 mycG - Micromonospora griseorubida  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Micromonospora griseorubida  
C:date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: S51594  
R:Inouye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S.

Mol. Gen. Genet. 245, 456-464, 1994  
A:Title: Characterization and expression of a p-450-like mycinamicin biosynthesis gene  
A:Reference number: S51593; MUID:95107242; PMID:7808395  
A:Accession: S51594  
A:Molecule type: DNA  
A:Residues: 1-397 <INO>  
A:Cross-references: EMBL:D16038; NID:g286050; PIDN:BAR03672.1; P'D:g303644  
C:Genetics:  
A:Gene: mycG  
A:Start codon: GTG  
C:Superfamily: Bacillus cytochrome P450 CYP206; cytochrome P450 homology  
C:Keywords: heme; oxidoreductase  
F:231-368/Domain: cytochrome P450 homology <CYP>

Query Match	28.3%;	Score 581;	DB 1;	Length 397;
Best Local Similarity	37.3%;	Pred. No. 9.3e-33;		
Matches 134;	Conservative	64;	Mismatches 131;	Indels 30;
				Gaps 8;

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: CB8170  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, K.R.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
Eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; PMID:21128732; PMID:11234002  
A;Accession: CB8170  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-434 <STO>  
A;Cross-references: GB:AL450380; NID:G33093682; PTDN:CA031043.1; GSPDB:GN00147  
C;Genetics:  
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

```

Query Match      28.3%; Score 581; DB 2; Length 434;
Best Local Similarity 35.1%; Pred. No. 1.le-32;
Matches 136; Conservative 80; Mismatches 140; Indels 32; Gaps 6;

QY      8 QALQRALINGKGNKODAVHPPEWYESRKQAPVSF--DEENQVWSVFLVDD--VKXAVGDKX 64
DB      35 QMLLLILIDPGTRAD--PPVTVRALIDYQPMQLPGMPLTVFSFSQDCEALRHPGLASD 91

QY      65 LPSSCMPQOOTSSIG-----NSLIIMDPKHITKIRSVVYNKAPTPTVMKQWEPRIOEIT 116
DB      92 RUKATIAQQAIAAGAEPRFPYASFSFPLDPDPDTRLRKLVSKAPAPKVVQVQALEGDIADLV 151

QY      117 DELIQFQGRSEFDLVHDSYPLPVIVISILLEGVSGNAHMEQFKAWSDLILVST-----168
DB      152 DSLIDKGAAGQDDVTDIADAPFLAVAVICKLLGVYVEDAFQFVRSALLVQSDVPFTTIT 211

QY      169 --PKDISEAEKAFLEERDKCEBELAAFFAGIIIEKKNKPEQDIIISILVFAEETSEKLSG 226
DB      212 GEPPATEERLRAGVWLRYLEQ-----LVKCRGTGPDGLISRLIELDESQDLITE 263

QY      227 ELIPIPTCLLLVAGNETTNLISNAMYSILETGCVTEELASHPELMPQAVEEARFRAPA 286
DB      264 EETIATCGLLVAGHETTNNLIANAVLAMPNFSQWKALSNNFQAPLVVVEETLYEYPAI 323

QY      287 PVLRRIAKRDTGIGGHLIKEGDMVLAVFASANDCEAKFDPRHMFDTJRRHNPNIHAFGHGI 346
DB      324 HLIGRVAAKDMVILGOTTILEGDTWVILLAAANDDPVAYSPPDFCDPDRPSRHLAFVGS 383

QY      347 HFCIGAPLARLEANIATLSLISAPFHEM 374

DB      384 HFCIGALARLEATVTLAISARFPQVQ 411

```

```

RESULT 11
  G83229      cytochrome P450 PA3331 [imported] - Pseudomonas aeruginosa (strain PA01);
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Feb-2001
C:Accession: G83229
A:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Kas, A.; Lathig, K.; Lim,
  ; Lory, S.; Olson, M.V.
  Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: G83229
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <STO>
A:Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AAG56719.1; SSCPDB:GN5C1
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3331
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
F:233-369/Domain: cytochrome P450 homology <P45>

```

RESULT 10

C37170  
probable cytochrome p450 [imported] - Mycobacterium leprae  
C3Species: Mycobacterium leprae

Query Match 27.8%; Score 571; DB 2; Length 418;  
 Best Local Similarity 34.1%; Pred. No. 4.9e-32;  
 Matches 124; Conservative 72; Mismatches 138; Indels 30; Gaps 6;

Qy 30 YESMRKDA--PVSFDEENQVSVFLYDVKVVGSD-----KELFSSCMPEQQTSS 76  
 Db 20 YDGLVDGAPRPAHRAAEHPVWVTVYDARKVLNHPGVVERDARQAAELVAKRTGSPAG 79  
 Qy 77 IGNSI-----INMOPKHTKIRSVVNKAFTRVWKQWMEPRIQEITDELIOKFGQSEPDLY 132  
 Db 80 IGEGLSHMLMLNDPPDTRLSRVLCRAFTPRQVERLQPHIERITAEALLDAMAGREQADLM 139  
 Qy 133 HDRSYPLVTVISLIGVPSAHMQPFK-ANSDLLVSTPKDSEAEKAFLEERDKCEEL 191  
 Db 140 ADFAPITAVIFELGIPAEAREHQS-----ERQALLSPPEEAQALADAQ 188  
 Qy 192 AAFPAGIIEKRKNKPEQDIISILVEASETEKLSGEBLIPFCTILLVAGNETTNLISNA 251  
 Db 189 VDYLRLVLEAKRPADDDVYSLVQADESQLSAEALVSMWHLMMSGFETTNMIGNA 248  
 Qy 252 MYSILETPGVYBELRSHPELMPQAVEALRFRPAPV-LRRIAKRDTTEIGHLIKEGDMV 310  
 Db 249 LVTLNVPEQLALIRAQPELPLWAMELVRHDSFVRASMLRFTVEDVELDGVITPAGEYI 308  
 Qy 311 LAFVASANRDEAKFDREHMDIRRHNPHTAFGHGHHFCLGAPLARLEANIALTSLISAF 370  
 Db 309 LVSNLTANHAERPDDELDRLDTRNTDGHLYGFGVHYCVGASLARDEGRITQRLIARF 368  
 Qy 371 PHME 374  
 Db 369 PDLQ 372

RESULT 12  
 B40634  
 N:erythromycin monooxygenase (EC 1.14.-.-) - Saccharopolyspora erythraea  
 N:Alternate names: erythromycin C-12 hydroxylase EryK  
 C:Species: Saccharopolyspora erythraea  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
 C:Accession: B40634  
 R:Stassi, D.; Donadio, S.; Staver, M.J.; Katz, L.  
 J. Bacteriol. 175, 182-189, 1993  
 A:Title: Identification of a Saccharopolyspora erythraea gene required for the final hydroxylation of erythromycin  
 A:Reference number: A40634; MUID:93106953; PMID:8416893  
 A:Accession: B40634  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-412 <STA>  
 A:Cross-references: GB:U82823; GB:L05776; NID:92327012  
 A:Note: sequence extracted from NCBI backbone (NCBIN:121243, NCBIPI:121245)  
 C:Genetics:  
 A:Gene: CYP113A1  
 C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
 F:238-376/Domain: cytochrome P450 homology <CYP>  
 F:354/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 27.6%; Score 565.5; DB 1; Length 412;  
 Best Local Similarity 39.1%; Pred. No. 1.2e-31;  
 Matches 136; Conservative 50; Mismatches 151; Indels 11; Gaps 6;

Qy 29 WYESMRKADAPVSPDEENQVSVFLYDVKVVGSDKELFSS---CMPQQTSSIGNSLINDM 85  
 Db 33 WLGTWEKQPV-WQDYGWVHVRHADVQTVLTDATFSSDFTRVIEGASPTPGMHEID 91  
 Qy 86 PPKHTKIRSVVNKAFTRVWKQWMEPRIQEITDELIOKFGQSEPDLYVHDPSPPLPVIVIS 145  
 Db 92 PPEHRLRKVSSAFPTRTISDLEPRIDRVTRSLA--DAGESFQVLDVLAFLPPTVIVA 149  
 Qy 146 ELLGVPSAHMQEOKANSOLLVSTPKDSEAEKAFLEERDKCEELAAFPAGIIEKRKN 205  
 Db 150 ELLGLPMDHQEGDWSGALVDIQMD--DPTDPALEARIDVNLPLTAYLKARCAERAD 207

```

Qy 206 PEQDIISILVEABETGEKLSSEELIPFCTLLILVAGNETTNLISNAMYSLLEIPGVVEEL 265
Db 208 PGDDLLISRLVLAEDVGRALDDEEAANFTALLAGHITTTVLLGNTVTRTDEHPAHDA 267
Qy 266 RSHPELMPQAVEELRRFAPAPVLRARIKRDTEIGGHLIKEGDMVLAFVASANDEAKFD 325
Db 268 AEDPARIPALVEEVLYRPPFPOMORTTTKATEVAGVPIPADVMVNTVWLSANRDSDAH 327
Qy 326 RPHMFD--IRRHNPHI--AFGHGTHFCGAPLARLEANIALTSLISAF 370
Db 328 DPDRFDPSAQVAPARTSSFGVGHVHFLAAPLARLENRVALEELIARF 375

RESULT 13
T30231
cytochrome P450 - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 15-Feb-2001
C:Accession: T30231
R:Apuricio, J.F.; Molnar, I.; Schweske, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A:Reference number: Z20782; MUID:96186896; PMID:8635756
A:Accession: T30231
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <APA>
A:Cross-references: EMBL:X86780; NID:G987088; PID:G987105; PIDN:CRA60465.1
C:Genetics:
A:Gene: rapN
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
P:238-375/Domain: cytochrome P450 homology <P45>

Query Match 27.5%; Score 564; DB 2; Length 404;
Best Local Similarity 34.9%; Pred. No. 1.4e-31;
Matches 130; Conservative 62; Mismatches 143; Indels 38; Gaps 7;

Qy 26 PFWYNSMRKDAPVSD-----EENQVSVFLYDDVKVKVGDKELFSSC-- 69
Db 14 PYPAEMERLEIHPYNRRLRAGELGRVIMLDGREAWLAASWEDVAKVFDPRFSSATL 73
Qy 70 ---MPQCTSSIGNS--IINMDPPKHTKRSVVKNAFTPRVMKQWEPRIQEIITDELIQKQ 124
Db 74 GKDVPRVLPAIQDPVIMLDPPEHTRLRVATKALTSSRMERALPRTEGVADLLDKML 133
Qy 125 GR-SEFDLVHDFSPLPVIVITSELNGVPSAHMEQPKAWSDDLIVSTPKDSEAEKAFLEE 183
Db 134 AKGAPADLMEDFALPLTIIICELLGVPIEDQTKFTWSDQMLSGAVSQEVVMAA---- 189
Qy 184 RDCCEEELAAFFAGLIEEKNKPKQDIISILVEABETGEKLSSEELIPFCTLLIVAGNET 243
Db 190 -----GQSLYVLSLIAERRKQDNNLLGSLVRARXKDDRLSETELVGFAVTLLAGYET 245
Qy 244 TTLNISNAMYSLLETPGVYSELRSHPPELMPQAVEALRRFAPAPVLR-----RIAKRDT 238
Db 246 TANAIGNSVYTLTHPEKLAELRKDLSLIPXAVDELLRI---IPIAKQASWTRMAVEDVE 302
Qy 299 IGGHLIKEGDMVLAFVASANDEAKFDRPHMFDIRRHNPHIAFGHGTHFCGAPLARLE 358
Db 303 LSGTIVKAGEAVATQTHSANTDPKVIDHPESIDFHRTSNPHMSLGHGAHHCMAQLVRVE 362
Qy 359 ANIATLSLISAEF 371
Db 363 MQTALGSLISRIP 375

RESULT 14
S71328
cytochrome P450 CYP219 - Sulfolobus solfataricus
N:Contains: oxidoreductase (EC 1.-.-)
C:Species: Sulfolobus solfataricus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000

```

C:Accession: S71328  
R:Wright, R.L.; Harris, K.; Solow, B.; White, R.H.; Kennelly, P.J.  
FBS Lett. 384, 235-239, 1996  
A:Title: Cloning of a potential cytochrome P450 from the Archaeon Sulfolobus solfataricus  
A:Reference number: S71328; MUID:96197795; PMID:8617361  
A:Accession: S71328  
A:Molecule type: DNA  
A:Residues: 1-368 <WRI>  
A:Cross-references: EMBL:U51337; NID:g1256447; PID:AAB03278.1; PID:g1256448  
A:Experimental source: ATCC 35091  
C:Genetics:  
A:Gene: CYP119  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:206-339/Domain: cytochrome P450 homology <CYP>  
F:317/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 26.7%; Score 548.5; DB 1; Length 368;  
Best Local Similarity 34.7%; Pred. No. 1.5e-30;  
Matches 134; Conservative 68; Mismatches 145; Indels 39; Gaps 10;

QY 27 PWYSEMRDAFVSDEENQVMSVFLYDDVKVVGDKELFSSCMP---QQTSSIGN---- 79  
Db 2 YDFSEMRKDPVYD--GNMQVFSYRYTKVNLNFSKFSDDLTCYHERLEDLRNGKIR 59  
QY 80 -----SIINDDPPKHTKIRSVNKAFTPRVMKQEPRIORITDELIOKFQGRSEFDLV 132  
Db 60 FDIPTRYMLTSDPLHDELRSMSADIFSPQKLQTLTFIRETRTSLSDSIDPR-EDDIV 118  
QY 133 HDFSPLPVIVISSELLGVPSAHMQEPAWSDILWSTPKDSE--EAKAFLEERDKCEE 190  
Db 119 KLIAPLPIIIVISKILSLPIEKEKEKENSILVAFLGPGGEIFELGCKYLE----- 170  
QY 191 LAAPFAGHIEKRNKPEQDIISLVEAETGKLSGEELIPFCTLLLVAGNETTTNLIN 250  
Db 171 ---LIGYVKDHLNS-GTEWSRVNVS-----NLSDIEKLGVIILLIAGNETTTNLIN 220  
QY 251 AMYSILETPGVVEELRSHPELMPQAVEALRPRAPVLRRAKTRTEIGGHLIKEGDMV 310  
Db 221 SVIDTRF- NLQRIREE-NLYLKAIEEALRSPVYMTVRKTRKERVKLGDQTIIEGYV 278  
QY 311 LAFVANSRDEAKRPHMFDIRRHNPPIAHGCHGHFCLGAPLARLEANIALTSLISAF 370  
Db 279 RWIANSRDEEWFHDGEKFIIDRPNPHLSFGSHLCLGAPLARLEARIIEEFSKRF 338  
QY 371 PHMECVSITPIENSIVYGLKSPRVKM 396  
Db 339 RHIEILDTEKVENELNGYKRLVRL 364

RESULT 15  
S18531  
Cytochrome P450 eryF - Saccharopolyspora erythraea  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Saccharopolyspora erythraea  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S18531; S16745  
R:Haydock, S.F.; Dowson, J.A.; Dhillion, N.; Roberts, G.A.; Cortes, J.; Leadlay, P.F.  
Mol. Gen. Genet. 230, 120-128, 1991  
A:Title: Cloning and sequence analysis of genes involved in erythromycin biosynthesis in methyltransferases.  
A:Reference number: S18530; MUID:92079886; PMID:1840640  
A:Accession: S18531  
A:Molecule type: DNA  
A:Residues: 1-406 <HAY>  
A:Cross-references: EMBL:X60379; NID:948941; PID:CAA42927.1; PID:g48943  
A:Note: the authors translated the codon AGG for residue 190 as Iys  
C:Genetics:  
A:Gene: eryF  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
F:238-375/Domain: cytochrome P450 homology <CYP>

Query Match 26.3%; Score 540; DB 1; Length 436;  
Best Local Similarity 33.7%; Pred. No. 6.6e-30;  
Matches 130; Conservative 72; Mismatches 132; Indels 52; Gaps 13;

QY 20 KQDAYHPPFPYVES---MRKDAPYS---FDEENQVMSVFLYDDVKVVGDKELFSSCMPQQ 73  
Db 8 ESQSFH-VDWRYTYAELRETAPVTVPVFLQDP-AWLVTGYDEAKAALSDRLSSD--PKK 63  
QY 74 -----TSSIGNSIINMDPPKHTKIRSVNKAFTPRVMKQEP 110  
Db 64 KYPQVEVERPVALGEPEDVENVFATNMGTS-----DPPTHRLKLVSQEFTVRVEARMP 119  
QY 111 RIGETIDELIOKFQGRSEFDLVHDFSYPPLPVIVISSELLGVPSAHMQEPAWSDILWSTP 169  
Db 120 RVEQITAEILLDREVGDGVVDIVDRFAHPLPIKVICELLGVDEKYRGFEFGWSSILWMDP 179  
QY 170 KDKSEEAFALEERDKCEEELAAFFAGIIEEKENKPEQDIISLVEAETGE-KLSGE 228  
Db 180 -----ERA--EORGQAAREVNFILDIIVERRRTERGDDLLSALIRVQDDDDGRLSADE 230  
QY 229 LIPFCTLLLVAGNETTTNLINAMYSILETPGVVEEL---RSHPELMPQAVEALRFRAP 285  
Db 231 LTSIALVLLLAGFEASVSLIGIGTYILLTHPD-QDQALVRRDPSALPNAVEELRYIAP 289  
QY 286 APVLRRIAKRDTIEIGHLIKEGDMVLAFAVANSRDEAKRPHMFDIRRHNPPIAHGCHG 345  
Db 290 PETTTFAAEVEIRGVVAIPQYSTVLVANGAANRDKQFPDPHFRFDVTRDTRGHLSPGQ 349  
QY 346 IHFCLGAPLARLEANIALTSLISAF 371  
Db 350 IHFCMRPLAKLEGEVALRALFORFP 375

Search completed: May 28, 2004, 23:23:41  
Job time : 72 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:10:41 ; Search time 41 Seconds  
(without alignments)  
502.921 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052  
Sequence: 1 MYVLRNALORALLNGKNK.....SITPIENSVIYGLKSRVNM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	100.0	396	1 YJTB_BACSU	O34374 bacillus su
2	839	40.9	405	1 CPXW_ARYOR	P27632 bacillus su
3	738	36.0	410	1 CPXW_ARYOR	O06069 bacillus me
4	711.5	34.7	410	1 CPXW_ARYOR	P14762 bacillus me
5	658	32.1	410	1 CPXW_ARYOR	O08469 bacillus su
6	646.5	31.5	405	1 CPXK_SACER	P33271 saccharopol
7	600	29.2	395	1 B101_BACSU	P53554 bacillus su
8	571	27.8	397	1 CPXQ_SACER	P48335 saccharopol
9	557	27.1	404	1 CPXK_SACER	Q00441 saccharopol
10	548.5	26.7	368	1 CPXW_SULSO	Q55080 sulfolobus
11	520	25.3	428	1 C124_MYCTU	Q50696 mycobacteri
12	516	25.1	402	1 C123_MYCTU	P77902 mycobacteri
13	514.5	25.1	398	1 C142_MYCTU	O53563 mycobacteri
14	499	24.3	433	1 C125_MYCTU	P71856 mycobacteri
15	489.5	23.9	402	1 C13B_XYLFA	Q99955 xyliella fas
16	488	23.8	414	1 C126_MYCTU	P77903 mycobacteri
17	486.5	23.7	402	1 C13B_XYLFT	O87ax5 xyliella fas
18	482	23.5	406	1 CPXQ_PSEPU	Q59723 pseudomonas
19	475	23.1	405	1 CPXQ_STRGO	P18326 streptomyce
20	474	23.1	400	1 C13C_XYLFA	Q99966 xyliella fas
21	471.5	23.0	438	1 C140_MYCTU	O08464 mycobacteri
22	465	22.7	399	1 C13C_XYLFT	O87av9 xyliella fas
23	462	22.5	402	1 CPXW_STRGO	P18327 streptomyce
24	451	22.0	405	1 C130_MYCTU	Q11062 mycobacteri
25	446	21.7	381	1 CPXG_STRSQ	P23296 streptomyce
26	439	21.4	409	1 CPXG_STRSQ	Q59831 streptomyce
27	437.5	21.3	402	1 NOR_FUSOX	P23295 fusarium ox
28	425	20.7	399	1 FAS1_RHOFA	P46373 rhodococcus
29	424	20.7	412	1 CPXW_STRGO	P26911 streptomyce
30	419.5	20.4	428	1 CPXK_PSEPU	P33006 pseudomonas
31	418.5	20.4	405	1 CPXW_BACSU	O34926 bacillus su
32	418.5	20.4	406	1 C5C4_ARYOR	Q8rn03 amycolatops
33	415	20.2	408	1 NOR2_CYLTO	Q12599 cylindrocac

34	403	19.6	403	1 NOR1_CYLTO	Q00616 cylindrocac
35	401.5	19.6	398	1 C5B3_ARYOR	Q8rn04 amycolatops
36	397	19.3	400	1 CPXP_RHISN	P55544 rhizobium s
37	391	19.1	401	1 CPXP_BRAJA	Q59203 bradyrhizob
38	390	19.0	436	1 THCB_RHOER	P43492 rhodococcus
39	385	18.8	489	1 C128_MYCTU	O59572 mycobacteri
40	376	18.3	400	1 C141_MYCTU	O08362 mycobacteri
41	375.5	18.3	429	1 CPXP_BRAJA	Q59204 bradyrhizob
42	366.5	17.9	414	1 Y4VG_RHISN	O53215 rhizobium s
43	362	17.6	414	1 CPXA_PSEPU	P00183 pseudomonas
44	355	17.3	391	1 C5A3_ARYOR	Q8rn05 amycolatops
45	351	17.1	422	1 CPXC_ARYOR	P24466 agrobacteri

## ALIGNMENTS

RESULT 1  
YJTB\_BACSU  
ID YJTB\_BACSU STANDARD; PRT; 396 AA.  
AC O34374;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative cytochrome P450 YJTB (EC 1.14.-.-).  
GN YJTB OR BSU12210.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Boletini A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerly I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
Entian K.D., Ewington J., Fabbri C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche R., Rose M., Sadate Y.,  
Sato T., Scanlan B., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tarakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambut S., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
"The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis".  
RL Nature 390:249-256 (1997).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF015825; AAC46317.1; --
CC DR EMBL; Z99110; CAB13078.1; --
CC DR PIR; B69851; B69851.
CC DR HSSP; Q55080; I109.
CC DR Subtilisin; BG13195; YjiB.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; p450; 2.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
CC Complete proteome. 349 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC FT METAL 349 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;
CC SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;

Query Match 100.0%; Score 2052; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.6e-124; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 0;

QY 1 MNVLRERQALQRALLGKKNQDAYHPPWYSEMRKDAVSPFDEENQVMSVFLYDDVKVW 60
DB 1 MNVLRERQALQRALLGKKNQDAYHPPWYSEMRKDAVSPFDEENQVMSVFLYDDVKVW 60

QY 61 GKELFSSCMPQOTSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOETIDELI 120
DB 61 GKELFSSCMPQOTSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOETIDELI 120

QY 121 QXFOGRSEFPLVDHFSVPLPVIIVISLGLVPSAHMEQFKAWSDLVSTPKDKSEAEKAF 180
DB 121 QXFOGRSEFPLVDHFSVPLPVIIVISLGLVPSAHMEQFKAWSDLVSTPKDKSEAEKAF 180

QY 181 LEERDKCEELAAFFAGIIEEKNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAG 240
DB 181 LEERDKCEELAAFFAGIIEEKNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAG 240

QY 241 NETTTLISNAMYSIILETPGVSELSHPELMPQAVEALRFPAPVLRRIAKRDEIG 300
DB 241 NETTTLISNAMYSIILETPGVSELSHPELMPQAVEALRFPAPVLRRIAKRDEIG 300

QY 301 GHLIKGDMVLAFVASANRDEAKFDRPHMEDIRRHNPHTAPGHGIFCLGAPLARLEAN 360
DB 301 GHLIKGDMVLAFVASANRDEAKFDRPHMEDIRRHNPHTAPGHGIFCLGAPLARLEAN 360

QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396
DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396

RESULT 2
CPXM_BACSU STANDARD; PRT; 405 AA.
AC P27632;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 109 (EC 1.14.-.-) (ORF405).
GN CYP109.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=W23;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;

```

```

RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL Gene 98:107-112(1991).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. They oxidize a variety of structurally unrelated
CC compounds, including steroids, fatty acids, and xenobiotics.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24523; AAA22720.1; --
CC DR HSSP; Q55080; I107.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; p450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Electron transport; Heme.
CC FT METAL 351 351 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 405 AA; 43845 MW; 1331D5BEA74B3C04 CRC64;

Query Match 40.9%; Score 839; DB 1; Length 405;
Best Local Similarity 43.7%; Pred. No. 1e-46;
Matches 167; Conservative 90; Mismatches 117; Indels 8; Gaps 2;

QY 19 NKQDAYHPPWYSEMRKDAVSPFDEENQVMSVFLYDDVKVVGDKELFSSCMPQOTSIG 78
DB 25 SEKDLFFPFPIYDKLRESEPVYDPLRDCWDVFKYDDVQFVKNPKLFSSKRGIT 80

QY 79 NSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOETIDELIOFGRSEFPLVDHFSY 138
DB 81 ESILTMPPKHTKIRSVVNKAFTPRVMKOWEPRIOETIDELIOFGRSEFPLVDHFSY 140

QY 139 LPVIVISELGLVPSAHMEQFKAWSDLVSTPKDKSEAEKAFLEERDKCEELAAFFAGI 198
DB 141 LPVIVISELGLVPSAHMEQFKAWSDLVSTPKDKSEAEKAFLEERDKCEELAAFFAGI 200

QY 199 IEKRNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAGNETTTLISNAMYSILET 258
DB 201 LSKERAPEKEDLMTLLQAHDIDGEYLTLEQIGFCILLVAGNETTTLISNAMYSILET 260

QY 259 PGVVEELRSHPELMPQAVEALRFPAPVLRRIAKRDEIGHIFCLGAPLARLEAN 318
DB 261 SVVQQQVRQNTDNNVANVIEETLRYSPVQAGRVAETEDTELGGVFIKKGSSVISWISAN 320

QY 319 RDEAKFDRPHMEDIRRHNPHTAPGHGIFCLGAPLARLEANALTSLISAPPHMECVS- 377
DB 321 RDEAKFDRPHMEDIRRHNPHTAPGHGIFCLGAPLARLEANALTSLISAPPHMECVS- 380

QY 378 ---ITPIENSVIYGLKSPRVKM 396
DB 381 DEKLEAIPSPFVFGVRLPVRI 402

RESULT 3
CPXM_BACME STANDARD; PRT; 410 AA.
AC Q06069;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 (MEG) (EC 1.14.99.-) (Steroid 15-beta-hydroxylase)
GN CYP106A2.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]

```



QY 127 SEFDVHDSYPLVIVISBELGVPSAHMQFKWASDLNVST-PKOKSEAEKAFLEERD 185  
 Db 135 STINIVDDLSPPFSLVIADLFGVPVDRYQFKKWDILFQPYDQERLEETE-----QEKQ 190  
 QY 186 KCEELAAFPAGLIBEKNKPEODIISILVAETGKLSGSELIPFCTILLVAGNETTT 245  
 Db 191 RAGAEFYQIDYPIVIEKNSLSDIISDLIOAEVDGTFDDEIVHATMLLGGAVETTS 250  
 QY 246 NLISNAMYSL-ETPGVYELSRHPELMPQAVEALFRAPAPVRLRIAKRDTGIGGHLI 304  
 Db 251 HALANFYSLYDDKSLYSELNRNRELAPRAVEMLRYRHSRDRKTVQDNELGVLK 310  
 QY 305 KEGDMVLAFVANSRDEAKTDRGEMFDIRHPN-PHIAFGHGHFCLGAPLARLENIAL 363  
 Db 311 KKGDDVVIAMNSACNMDETWFENPFSVDIHRPTNKKHJLTFGNGPHFCLGAPLARLEMKIIL 370  
 QY 364 TSLISAPPHMECVSITPIEN 383  
 Db 371 EAFLEAFSHIE-----PFED 385

RESULT 5  
 CPXY BACSU STANDARD; PRT; 410 AA.  
 ID CPXY BACSU STANDARD; PRT; 410 AA.  
 AC O08469;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome P450 (BC 1.14.-.-).  
 GN CYPA OR CYP107J1 OR BSU26740.  
 OS Bacillus subtilis  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=97431495; PubMed=9287000;  
 RA Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;  
 RT "An lrp-like gene of Bacillus subtilis involved in branched-chain  
 amino acid transport.";  
 RL J. Bacteriol. 179:5448-5457 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97453479; PubMed=9308178;  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 RA Duesterhoeft A., Ehrlich S.D.;  
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of  
 the lev operon reveals two new extracytoplasmic function RNA  
 polymerase sigma factors, Sigv and Sigz.";  
 RL Microbiology 143:2939-2943 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Mozer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetel D., Porwollik S., Prescott A.M.,

RA Presecan E., Putic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpatra P., Tognoni A.,  
 RA Teotato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler Z., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Darchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RL Nature 390:249-256 (1997).  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR ENBL; Y11043; CAA71937.1; -;  
 DR ENBL; U93876; AAB80898.1; -;  
 DR ENBL; Z99117; CAB14615.1; -;  
 DR PIR; E69611; E69611.  
 DR HSP; Q03441; I0XA.  
 DR Subtilisin; BGL1929; CYP4.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME P450; 1.  
 DR Oxidoreductase; Monooxygenase; Heme; Complete proteome.  
 KM METAL 359 359 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 410 AA; 47384 MW; 035E9858A4761AB CRC64;

Query Match 32.1%; Score 658; DB 1; Length 410;  
 Best Local Similarity 37.4%; Pred. No. 3.9e-35;  
 Matches 146; Conservative 78; Mismatches 136; Indels 30; Gaps 8;

QY 1 MVLINPRQALQRLNGKQDAYHFFPWYESMRKDA---PVSFDEENQVWSVFLYDDWK 57  
 Db 9 VTILTESQLSSRAF-----KDEAYE---FYKELRKQALYPLSLGALGKGLISRYDDAI 60  
 QY 58 KVVGDKEI-----PSS-----CMPQQTSSIGNSIINMDPPKHTKIRSVNKAFTPRVM 105  
 Db 61 HLLAKNEKLKNYENVETAKEKPAALLKNEETLYKMNSDPPDNNRRLTLVQKAFTHMI 120  
 QY 106 KQWEPRIQBITDELQKQFQSEFIDLVDHDFSVPLPVIVISBELGVPSAHMQFKWASDL 165  
 Db 121 LQLEDKIQHIADSLDDKVPKFMNLVDVYAPFLPIIVISEXMLGIPLEDKQKFRVWSQAI 180  
 QY 166 VSTPKDKSEAEKAFLEERDKCEELAAFFAGLIEKRNKPEQDIIISILVEABETEGLKS 225  
 Db 181 I-----DFSDAPER--LQENDHLLGEFVEYLSLVKRRREPAGDLISALIOASEGTQLS 234  
 QY 226 GBELIPFCTLLVAGNETTLLISNAMYSLTPTPGVYELSRHPELMPQAVEALFRAP 285  
 Db 235 TSELVSMMLLVAGHETTVNLTNTWYALMCHHCHQLEKLRQCPDLMNSAIEALRHPSP 294  
 QY 286 APVLR-RIAKRDTGIGGHLIKEGDMVLAFVANSRDEAKFDRPHMFDIRHPNPHIAFGH 344  
 Db 295 VELTTIRWTABPFLHGOEIKRKQDVIIISANRDEKIFPNADIPDIERKNRHIAFGH 354  
 QY 345 GHFCLGAPLARLENIALTSLISAPPEME 374  
 Db 355 GNHFCLGAPLARLENIALTSLILRRCPNIQ 384

RESULT 6  
 CPXY SACER  
 ID CPXY SACER STANDARD; PRT; 405 AA.



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC CC
CC EMBL; U51868; AAB17462.1; -.
CC EMBL; AF008226; AAC00266.1; -.
CC EMBL; Z99119; CAB14997.1; -.
CC DR PIR; G69594; G69594.
CC DR HSP; Q55080; 1107.
CC DR Subtilist; BG11528; biof.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; Cytochrome_P450; 1.
CC KW Biotin biosynthesis; Oxidoreductase; Monooxygenase;
CC Electron transport; Heme; Complete proteome.
CC METAL 345 345 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE 395 AA; 44865 MW; E4AC3AF2637ACE1A CRC64;
CC
CC Query Match 29.2%; Score 600; DB 1; Length 395;
CC Best Local Similarity 35.5%; Pred. No. 1.9e-31;
CC Matches 128; Conservative 84; Mismatches 131; Indels 18; Gaps 7;
CC
CC QY 25 HPFPKYESNRKDAV---SFDEENQVWSVELYDDVKVGVGDKEL-----FSSCMPOQT 74
CC DB 15 NPSFYDTLRVHPIYKGSF-LKPGWVVTGHEETAALKDARFKURTELPSSSTKYQDL 73
CC
CC QY 75 SSGNS-IINMDPPKTKIRSVNKAFTPRVMKQWEPRIQETIDELIQKQGRSEFDLVH 133
CC DB 74 SHVQNMWLFQNPQDHRRLRTLASGAFRTTESYQYIIETVHLLDQVQKKRMEVIS 133
CC
CC QY 134 DFSYPLPVIVISLGVSAHMEQKANSLLVST-PRDKSEAKAFLEERDKCEELA 192
CC DB 134 DFAPPLASVIANIIGVPEDEQKWAASLIQITDFTRSKA----LTEGNIMAVQAM 189
CC
CC QY 193 AFEAGLIEEKNKPPQDIISILVEAETGKLSGELIPFCITLLVAGNETTINLSNAM 252
CC DB 190 AFFKELIQKRPQDQWISMLKREK-DKLTREAASTCILLAGHETTIVNLSNV 248
CC
CC QY 253 YSILETPGVYELRSHPELMPQAVEALFRAPAVLRRIRAKRDTTEIGGHLIKEGDMVLA 312
CC DB 249 LCLLQHPQLKRLNPDLIGTAVECLRYESPTQMTARVASEDIDCGVTIRQGEQVYL 308
CC
CC QY 313 FVASNRDEAKFDRHMDIRHNPHTAFGHIHFCGAPLARLENNIALTSLISAFPH 372
CC DB 309 LLGAANRDEPSTFNPVDFITRSPNPHLSFGHHVCLGSSSLARUEAQIAINTLQRMPS 368
CC
CC QY 373 M 373
CC DB 369 L 369
CC
CC RESULT 8
CC CPXQ_SACER STANDARD; PRT; 397 AA.
CC AC P48635; O33990;
CC DT 01-FEB-1996 (Rel. 33, Created)
CC DT 15-FEB-1998 (Rel. 37, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Cytochrome P450 113A1 (Erythromycin B/D C-12 hydroxylase).
CC GN ERYK OR CYP113A1
CC OS Saccharopolyspora erythraea (Streptomyces erythraeus).
CC OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC OC Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
CC OX NCBI_TaxID=1836;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-NRRL 2338;
CC RX MEDLINE=97390129; PubMed=9249068;
CC RA Pereda A., Summers R., Katz L.;
CC RA "Nucleotide sequence of the ermE distal flank of the erythromycin
CC RT biosynthesis cluster in Saccharopolyspora erythraea.";
CC DE 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE 6-deoxyerythronolide B hydroxylase (EC 1.-.-.-) (6-DEB hydroxylase)
CC DE (Erythromycin A biosynthesis hydrolase) (Cytochrome P450 167A1)
CC DE (CYPCV1A1) (P450eryf).
CC RL J. Bacteriol. 175:182-189 (1993).
CC RN [2]
CC RP REVISIONS.

```

```

RC STRAIN-NRRL 2338;
RC MEDLINE=97390129; PubMed=9249068;
RC RA Pereda A., Summers R., Katz L.;
RC RT "Nucleotide sequence of the ermE distal flank of the erythromycin
RC RT biosynthesis cluster in Saccharopolyspora erythraea.";
RC RL Gene 193:65-71 (1997).
CC -!- FUNCTION: Responsible for the C-12 hydroxylation of the
CC macrolactone ring of erythromycin.
CC -!- PATHWAY: Erythromycin B and D biosynthesis; final hydroxylation
CC step.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U82823; AAC45584.1; -.
CC PIR; B40634; B40634.
CC HSP; Q55080; 1107.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; Cytochrome_P450; 1.
CC Oxidoreductase; Monooxygenase; Electron transport; Heme;
CC Antibiotic biosynthesis.
CC METAL 339 339 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE 397 AA; 43759 MW; 8306AF79414C981E CRC64;
CC
CC Query Match 27.8%; Score 571; DB 1; Length 397;
CC Best Local Similarity 39.2%; Pred. No. 1.4e-29;
CC Matches 136; Conservative 49; Mismatches 152; Indels 10; Gaps 5;
CC
CC QY 29 WTESNRKDAVSEFDEENQVWSVELYDDVKVGVGDKELFSS---CMPQQTSSGNSIINMD 85
CC DB 19 WLGTRKQPV-WQDRYGVWVFRHADVQTVLRDTATFSSDPRVIEGASPTPGMTHIED 77
CC
CC QY 86 PKKTKIRSVNKAFTPRVMKQWEPRIQETIDELIQKQGRSEFDLVHDFSYPLPVIVIS 145
CC DB 78 PEHRAIRKVSASATPTISDLPRIDVTRSLA--DAGESFDLVDLVLAFLPPTIVA 135
CC
CC QY 146 ELGVPSAHMEQKANSLLVSTPKDSEAKAFLEERDKCEELAFAFAGIEEKRNK 235
CC DB 136 ELGLPXPXDEHQFGWSGALVDIQMD--DPTPALAERIAVLNPLTAYLKARCAERRAD 193
CC
CC QY 206 PQDIIISILVEAETGKLSGELIPFCTLLVAGNETTINLSNAMYSILETPGVYVEL 265
CC DB 194 PGDDLISRLVLAEDVGRALDDEEANEFTALLAGHITTVLGNIVRTILDEHPAHDAA 253
CC
CC QY 266 RSHPELMPQAVEALFRAPAVLRRIRAKRDTTEIGGHLIKEGDMVLAFLVASNRDEAKFD 325
CC DB 254 AEDPGRIPAIVEVLRYRPPQMQKRTTKATEVAGVPIPAQVMVNTWLSNRNDSADHD 313
CC
CC QY 326 RPHMFDIRHNP--PHIAPGHGHPCLGAPLARLENNIALTSLISAF 370
CC DB 314 DPDRFDPKSGGAAQSFSGHGVHFCGAPLARLENRVALEEIIARF 360
CC
CC RESULT 9
CC CPXJ_SACER STANDARD; PRT; 404 AA.
CC ID CPXJ_SACER
CC AC Q00441;
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE 6-deoxyerythronolide B hydroxylase (EC 1.-.-.-) (6-DEB hydroxylase)
CC DE (Erythromycin A biosynthesis hydrolase) (Cytochrome P450 167A1)
CC DE (CYPCV1A1) (P450eryf).
CC GN ERYF OR CYP107A1.
CC OS Saccharopolyspora erythraea (Streptomyces erythraeus).

```

```
Query Match      27.1%; Score 557; DB 1; Length 404;
Best Local Similarity 33.9%; Pred.No. 1.le-28;
Matches 130; Conservative 71; Mismatches 134; Indels 48; Gaps 11;
```

Db 8 ESDSPH-VDMYNTYAEALRETAFTVTRFUGQD-AWLVTGYDEAKAALSLRLSSD--PKK 63  
 QY 74 -----TSSIGNSIINMPDPKHTKIRSVVNVKAFTRPVAKQWEP 110  
 Db 64 KYPGVEVEFPALGPEOVNRYFAINMGTS-----DPPTHRLKLVQSEFTVRRVEAMP 119  
 QY 111 RIQETIDELIQKQGRSEFDLWHDPSYLPVIVISSELLGVPSAHMEQFAM-SDLLVSTP 169  
 Db 120 RVEQITABELLDEWGSVVDIVDRFAHPILPIKVICELLGVDBKRGFGEFRWSEILVMDP 179  
 QY 170 KDKSEAKAFLEERDKCEELAAFPAGIIEKRNKPEODIISILVEAETGE-KLSGEE 228  
 Db 180 -----BRA--EQRQQAAREVNFILDLVERRRTEPGDLSALIRVQDDDDGRLSADE 230  
 QY 229 LIPFCTLLVAGNETTNLISAMYSILETPGYEELRSHPELMPQAVREALFRAPAFV 288  
 Db 231 LTSIALVLLAGFEASVSLIGITGYLLTHPQALVRRDPSALPNAVEILRYAPPET 290  
 QY 289 LRIAKRDTIEGHLIKEDGMVLAFAVANRDEAKFDRHMDIRHPNPHAFAGHGHF 348  
 Db 291 TTRFAAEVEIGGVAIPQYSTVLVANGAANRDPKQFPDRFDVTRDTRGHLSEFGQGIHF 350  
 QY 349 CUGAPLAEALNIALTSLISAPP 371  
 Db 351 CMGRPLAKLGEVAURLAFGRFP 373

## RESULT 10

CPXW SULSO  
 ID CPXW SULSO STANDARD; PRT; 368 AA.  
 AC Q5080;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome P450 119 (EC 1.14.14.-).  
 GN CYP119.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35091 / DSM 1616 / P1;  
 RA MEDLINE=96197795; PubMed=8617361;  
 RX Wright R.L., Harris K., Solow B., White R.H., Kennelly R.H.;  
 RT "Cloning of a potential cytochrome P450 from the archaeon Sulfolobus solfataricus";  
 RL FEBS Lett. 384:235-239 (1996).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=99032823; PubMed=9813164;  
 RA McLean M.A., Maves S.A., Weiss K.E., Krepich S., Sligar S.G.;  
 RT "Characterization of a cytochrome P450 from the acidothermophilic archaea Sulfolobus solfataricus";  
 RL Biochem. Biophys. Res. Commun. 252:166-172 (1998).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20261529; PubMed=10799487;  
 RA Koo L.S., Tschirret-Guth R.A., Straub W.E., Moenne-Loccoz P.,  
 RA Loehr T.M., Ortiz de Montellano P.R.;  
 RT "The active site of the thermophilic CYP119 from Sulfolobus solfataricus";  
 RL J. Biol. Chem. 275:14112-14123 (2000).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RX MEDLINE=20414776; PubMed=10957637;  
 RA Park S.-Y., Yamane K., Adachi S.-I., Shiro Y., Weiss K.E.,  
 RA Sligar S.G.;  
 RT "Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from Sulfolobus solfataricus";  
 RL Acta Crystallogr. D 56:1173-1175 (2000).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.93 ANGSTROMS).

RX MEDLINE=20469464; PubMed=10859321;  
 RA Yano J.K., Koo L.S., Schuller D.J., Li H., Ortiz de Montellano P.R.,  
 RA Poulos T.L.;  
 RT "Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus";  
 RL J. Biol. Chem. 275:31086-31092 (2000).  
 CC -!- FUNCTION: CATALYZES THE H(2)O(2)-DEPENDENT EPOKIDATION OF STYRENE, CIS-BETA-METHYLSTYRENE, AND CIS-STILBENE WITH RETENTION OF STEREOCHEMISTRY.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- MISCELLANEOUS: Stable at up to 85 degrees Celsius.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC -!- CAUTION: DOES NOT SEEM TO EXIST IN THE COMPLETE GENOME OF P2 STRAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U51337; AA030278.1; -;  
 DR F1R; S71328; S71328.  
 DR PDB; 1F4T; 23-OCT-00.  
 DR PDB; 1F4U; 23-OCT-00.  
 DR PDB; 1IO7; 21-JAN-03.  
 DR PDB; 1IO8; 21-JAN-03.  
 DR PDB; 1IO9; 21-JAN-03.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME P450; 1.  
 KW Oxidoreductase; Monooxygenase; Electron transport; Iron; Heme;  
 KW 3D-structure.  
 FT ACT\_SITE 213 213  
 FT METAL 317 317 IRON (HEME AXIAL LIGAND).  
 FT HELIX 2 11  
 FT STRAND 14 16  
 FT STRAND 21 23  
 FT HELIX 25 34  
 FT TURN 36 38  
 FT STRAND 39 40  
 FT HELIX 46 53  
 FT TURN 54 56  
 FT HELIX 63 65  
 FT HELIX 68 70  
 FT TURN 73 74  
 FT HELIX 75 80  
 FT HELIX 81 83  
 FT TURN 84 87  
 FT HELIX 89 108  
 FT TURN 109 109  
 FT TURN 112 113  
 FT STRAND 115 116  
 FT HELIX 117 123  
 FT TURN 121 123  
 FT HELIX 124 134  
 FT TURN 135 135  
 FT TURN 138 143  
 FT HELIX 141 147  
 FT HELIX 148 150  
 FT TURN 152 153  
 FT TURN 158 159  
 FT HELIX 161 178  
 FT TURN 179 180  
 FT HELIX 184 190  
 FT TURN 191 191  
 FT HELIX 196 208  
 FT TURN 209 210  
 FT HELIX 211 227  
 FT TURN 228 229

```

FT 231 237
FT 238 238
FT 240 250
FT 254 254
FT 254 254
FT 257 261
FT 267 267
FT 268 269
FT 270 272
FT 274 275
FT 277 280
FT 282 285
FT 286 287
FT 289 291
FT 293 296
FT 300 301
FT 309 310
FT 313 315
FT 318 319
FT 320 335
FT 336 337
FT 340 348
FT 354 354
FT 357 365
SQ SEQUENCE 368 AA; 42863 MW; E2A5C4F064537E8C CRC64;

Query Match      26.7%; Score 548.5; DB 1; Length 368;
Best Local Similarity 34.7%; Pred. No. 3.4e-28;
Matches 134; Conservative 68; Mismatches 145; Indels 39; Gaps 10;

QY 27 PPWYSEMRKADPVGFDEENQVWSVFLYDDVKKVGVGDKELFSQMP---QQSSSTGN----- 79
Db 2 YDWFSEMRKKDPVYD--GNWQVFSYRYTKVELNLFNFKSFDLTGTYHERLEDLNGKIR 59
QY 80 -----SIINMDPPKTKIRSVNKAFTPRVMKOWEPRIQETIDELIKQFGSERFDLV 132
Db 60 FDIPTRYMTSDPLDELHLSMGADIFSPQKQTLFTFIRETRSLDSDIDPR-EDDIV 118
QY 133 HDFSYPFLPVISILLGVPSAHMEQFKAWSLLVSTPKDKSE--EAEKAFLEDRDKCEE 190
Db 119 KKLAVPPIIIVISKILGLPIEDKEKFKWSDLVAFRLGKPGCEIFELGKKYLE----- 170
QY 191 LAFFAGIIEBKRNKPDQDIISILVEAEETGKLSGBELPFCILLVAGNETTNLISN 250
Db 171 -----LIGYVKDHLNS-GTEVVSRYVNS-----NLSDIEKLGYYIILLIAGNETTNLISN 220
QY 251 AMYSILETPGVSELRGHPMLPOAVEALRFAPAPVLRRIAKRDTGIGHLIKGDMV 310
Db 221 SVIDFTRF-NWQRIREE-NLYLKAIEALRYSPVMTVTKERVKLGQTLSESYV 278
QY 311 LAFVANSRDEAKFDRPHMFDIRHPNPHIAPGHIHFCIGAPLARLEANTALSLISAF 370
Db 279 RVWIASANRDEEVEFDGKEKTPDRNPPLHSFGSGIHLCLGAPLARLEARIAIEFBSKRF 338
QY 371 PHMECVSITPIENSIVYGLKSFYKM 396
Db 339 RHIEILDTEKVPNEVLNGYKRLVRL 364

RESULT 11
C124_MYCTU
ID C124_MYCTU STANDARD; PRT; 428 AA.
AC Q50696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 124 (EC 1.1.1.4.-.-).
GN CYP124 OR RV2266 OR MT2328 OR MTCY339.44C OR MB2289.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN 1:

```

```

RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva V., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z77153; CAB00969.1; -.
CC EMBL; AE007076; AAK46610.1; -.
CC EMBL; BX248342; CAD97150.1; -.
CC PIR; F70729; F70729.
CC HSSP; P33006; 1CPT.
CC TIGR; MT2328; -.
CC TubercuList; Rv2266; -.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PROSITE; PS00086; CYTOCHROME_P450; FALSE NEG.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
CC Complete proteome.
FT METAL 379 379 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 23 23 D -> G (IN REF. 2).
FT CONFLICT 75 75 Y -> N (IN REF. 2).
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3CSAE348591 CRC64;

Query Match      25.3%; Score 520; DB 1; Length 428;
Best Local Similarity 32.1%; Pred. No. 2.7e-26;
Matches 126; Conservative 82; Mismatches 148; Indels 36; Gaps 10;

QY 30 YESMRKADPV-----EDEENQVWSVFLYDDVKKVGVGDKELFS-----CMFQQTSS 76
Db 43 FATLRREAPISFWFTIELPGFVAGNHWALTKYDDVFYASRHPDIFSSYPNTINDQPE 102
QY 77 IG---NSTINMDPPKTKIRSVNKAFTPRVMKOWEPRIQETIDELIKQFGSR---EFD 130
RN 1:

```



Db 103 LAEYFGSMVLDDPHQRLRSIVSRAFTKVVAREAAVRAHRLVSSMIANRPDQAD 162  
 Qy 131 LVHDSFVPLVIVISSELLGVSAHMEQKAWSDLLVS-TPDKSREAKAFLEERDKCEE 189  
 Db 163 LVSELAGPLPQIICDMGIPKADHQRIHFHWINVLFGDPLAD-----FDFMQVSA 217  
 Qy 190 ELAAFFAGIIEKENKPEQDIISILVEAEETGKLSGELIPFCTLLVAGNETTINLIS 249  
 Db 218 DIGAYATAEDRRVNHDDLTSSIVAEVDGERLSREIASFFLLVAGNETTRNAT 277  
 Qy 250 NAMYSIETGCVYBELSH-----PELMPQAVEARLPRAPVLRRIAKRDTGIGHLI 304  
 Db 278 HGVLALSR-----YPEQRDRWSDDEGLAPTAVEEIVRWASFVWYMRRTLTQDIELGTGM 333  
 Qy 305 KEGDMVLAVASANDRAKDRPHMDFIRRNPHIAP-GEIHFCIGAPLARLEANIAL 363  
 Db 334 AAGDKVLSWYCSANDRSKFDADPTWTFDARPNPHLFGGGAHFCIGANLARRIRVAF 393  
 Qy 364 TSLISAPFMBCV-SITPIENSVIYGLKSRV 394  
 Db 394 DELRRQMDVVATEEPARLLSQFINGIKTLFV 425

RESULT 12  
 ID C123 MYCTU STANDARD; PRT; 402 AA.  
 AC P7792;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative cytochrome P450 123 (EC 1.14.-.-).  
 GN CYP123 OR RV0766C OR MT0790 OR MTCY363.11C OR MB0789C.  
 OS Mycobacterium tuberculosis, and  
 OS Mycobacterium bovis  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773, 1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;  
 RX MEDLINE=92206494; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,  
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 Deicher A., Uitterlbeck T., Weidman J., Khouri H., Gill J., Mikula A.,  
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."

RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; Z80226; CAB02396.1; -.  
 DR EMBL; AE006970; AAK45032.1; -.  
 DR EMBL; BX248336; CAD3651.1; -.  
 DR PIR; A70707; A70707.  
 DR HSSP; Q00441; 10XA.  
 DR TIGR; MT0790; -.  
 DR TubercuList; RV0766C; -.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 KW Complete proteome.  
 FT METAL 350 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 402 AA; 45421 MW; 76AD8430.9361798 CRC64;  
 Query Match 25.1%; Score 516; DB 1; Length 402;  
 Best Local Similarity 33.5%; Pred No. 4,6e-26;  
 Matches 120; Conservative 72; Mismatches 152; Indels 14; Gaps 4;  
 QY 26 PFPWYESMRKDAVPSFDEENQVSVLYDYDKVGVGDKELFS-----SCMPQQTSSIG-- 78  
 Db 23 PYPYVRLRDEAPLYRNEERNFVAVSRHVDLQGPDRDSTALSNAYGVSLDPSRSTSEAYR 82  
 QY 79 -NSIINMDPPKTKLRVYNKAFTRVNMKQEPRIQETIDELIQFQGSSEFDLVHDSY 137  
 Db 83 VMSMLAMDPPAHLRMTILVSKGFTPRIRIE-EPQVLELARLHLSALQTESDFVAEFAG 142  
 QY 138 PLPVIVISELLGVSAHMEQKAWSDLLVSTPKDSEAEKAFLEERDKCEEELAAFFAG 197  
 Db 143 KLPMDVISELIGVPTDTRARFALADAVLHREDGVADVPPFAX-----AASIELMRYVAD 197  
 QY 198 IIEEKNRPEQIISILVEAEETGKLSGELIPFCTLLVAGNETTINLISNAMYSLILE 257  
 Db 198 LIAEFRFRPNANLTSALLAAELDGRLSQDEIMAFELMWIAGNETTTTLLANAVYWAH 257  
 QY 258 TPGVYEELRSHPELMPQAVEARLPRAPVLRRIAKRDTGIGHLIKEGDMVLAFVSA 317  
 Db 258 HPGQLARVFDHRSRIPMWVETLRYDTSSQILARTVAHDLTYDTTPEGEVILLPGSA 317  
 QY 318 NRDEAKFRPHMFDIRRHFN-PIAFGEGIHFCIGAPLARLEANIALTSLISAFPHME 374  
 Db 318 NEDRVFDDPDYRIGREIGCKLVSGSGAHCIGLGHARMEARVALGALLRRIENYE 375  
 RESULT 13  
 ID C142 MYCTU STANDARD; PRT; 398 AA.  
 AC O53563;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative cytochrome P450 142 (EC 1.14.-.-).  
 GN CYP142 OR RV3518C OR MT3619 OR MTV023.25C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon K., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL022022; CAAL7755.1; -;  
CC EMBL; AE007165; AAK47979.1; ALT\_INIT.  
CC HSSP; P33006; LCPT.  
CC TIGR; MT3619; -;  
CC Tuberculist; RV3518c; -;  
CC InterPro; IPR001128; Cytochrome\_P450.  
CC Pfam; PF00067; P450; 1.  
CC PRINTS; PR00385; P450.  
CC PROSITE; PS00086; CYTOCHROME\_P450; 1.  
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
CC Complete proteome. 340 IRON (HEME AXIAL LIGAND) (BY SIMILARITY)  
CC METAL 340 MTEAPVDVLAQGNFYASR -> GMSPNRSRAPPSRRP  
CC CONFLICT 1 18 {IN REF. 20}  
CC SEQUENCE 398 AA; 44398 MW; BCF3C23ECB5767F CRC64;  
CC  
CC Query Match 25.1%; Score 514.5; DB 1; Length 398;  
CC Best Local Similarity 34.2%; Pred. No. 5.6e-26;  
CC Matches 129; Conservative 61; Mismatches 164; Indels 23; Gaps 7;  
CC  
CC QY 30 YESRKDAPVGFDERNOVMSVFLYDDVKKVGDKEFLFSSC---MPQQTSSIGNSIINMDP 86  
CC DB 24 YRWRANQPV-FRDRNGLAASSTYQAVIDAEPQLFNSNAGGIRPDQPAL--PMMLMD 80  
CC QY 87 PKHTKIRSVNKAFTPRVWKQWEPRIQETDLELOKFGQSRFDFLVDHPSYLPVIVISE 146  
CC DB 81 PAHLRLRLVNAFTGTRKRVKDEASIAALCDTLIDAVCEGRCDFVRDLAAPLMAVIGD 140  
CC QY 147 LIGVSAHMEQKANSLLVLS-----TPKDKSEAEKAFLEERDKCEELAAFFACIIE 201  
CC DB 141 MLGVPRQEDMLRWSDLLWTLFSSVSGEDFQITMDFAAYND-----FTRATIAA 192  
CC QY 202 KENKQEQDIISILVRAEETGEKLSGELIPFCTLLVAGNETTTLNISNAMYSILETGPV 261  
CC DB 193 READPTDULSVLVSEVNGERLSDELVWETLLIIGGDETRHTLSGTEQLLRNQ 252  
CC QY 262 YEELRSHPELMPQAVEALFRAPAPVLRARIKROETIGHLIKEDMVLAFVANSRDE 321  
CC DB 253 WDLQDPDLSLFGAIEELRWTAFTVKNMCKVLTDTEFFGTALCAGEKNMLPFESANFDE 312

QY 322 AKFDRPHMDIRRHDPHIAFGHIGHCGLAPLARLEANIALTSLSAFPHMECV---SI 378  
DB 313 AVFCEPEKEDVQVQNPNSHLAFGFGTHFCGNQLARLELSMTXERVLRPLDLRLVADSV 372  
QY 379 TPTE-NSVIYGLKSRV 394  
DB 373 LPLRPANFVSGLESMEPV 383  
  

RESULT 14  
C125\_MYCTU STANDARD; PRT; 433 AA.  
ID \_C125\_MYCTU  
AC P71856;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative cytochrome P450 125 (EC 1.14.-.-).  
GN CYP125 OR RV3545C OR MT3649 OR MTCY03C7.11 OR MB3575C.  
OS Mycobacterium tuberculosis, and  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
RX MEDLINE=9634230; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon K., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.bovis; STRAIN=AF2122/97;  
RX MEDLINE=22709107; PubMed=12788972;  
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,  
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RA "The complete genome sequence of Mycobacterium bovis.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
RL -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z82098; CAB05061.1; -;  
CC EMBL; AE007166; AAK48008.1; -;

EMBL: BX248346; CAD95761.1; --  
 PIR: B70677; B70677.  
 HSP: P33006; 1CPT.  
 TIGR: M3649; --  
 DR Tuberculin; RV3545c; --  
 DR Cytochrome\_P450.  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 2.  
 DR PROSITE: PS00086; CYTOCHROME P450; PALSE NEG.  
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 DR Hypothetical proteome. 377 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 DR METAL 377  
 DR SEQUENCE 433 AA; 48432 MW; 48432 MW; FCBE0594D92171P4 CRC64;  
 Query Match 24.3%; Score 499; DB 1; Length 433;  
 Best Local Similarity 30.6%; Pred. No. 6.1e-25;  
 Matches 121; Conservative 80; Mismatches 157; Indels 38; Gaps 10;  
 CC 26 PFWYVESMRKADPVSD-----EENQVWGVFLYDDVKKVVGKELFSS-----CMP 71  
 CC 39 PVAEFABLSAIPWNGQDPKGGGHDGFGFWAITKLVNDVKEISRHSDVFSYVGVIP 98  
 CC 72 QQTSSIGNS-----IINMDPPKHTKIRSVNKAFTPRVWKQWEPRIQELITELIOKF 123  
 CC 99 RFKNDIAREDEVQFVWMLNDAPHTLRLKIIISRGFTFRAVGLHDELQBRQAIAEA 158  
 CC 124 QSRSEFOLVHDFVPLPVIIVISLIGVSAHMEQPKAMSDLLVSTPKDSBEAEAFLEE 183  
 CC 159 AAGSGDEVEQVSCPLQAIAGLGVFOEDRGKLFHNSMTG-----NEDEYAHIDP 213  
 CC 184 RDKCEEELAAFPAGIIEKRNKPEQDIISILVEABETGEKLSGELIPCTLLLVAGNET 243  
 CC 214 K-ASSAELIGVAMKAEKAKNPADDIVTLQADIDGKLSDDDFGFWFVWMLAVAGNET 272  
 CC 244 TTNLSINAMYSILETPG---VYEELRSHPELMPQAVEALRFRAPVLRRAKRDTEIG 300  
 CC 273 TNSITQGMVAEHPDQWELYKKVR--PE---TAADEIVRWATPVTAFQRLRDYELS 327  
 CC 301 GHLIKEGDMVLAFVASANDEAKFDRPHMFDIRRHNPPIAF--GHGIFHCLGAPLARLEA 359  
 CC 328 GVQIKKGQVWVWFYSANFDEEVQDFPFNIDLRNPNHVGFGTGAHYCIGANTARMTI 387  
 CC 360 NIALISLISAFHMECVSITP--IENSIVIVGLKSFV 394  
 CC 388 NLIFNAVADHMDPLKISAPERLSGLWGLKHWQV 423  
 DR Krieger J.E., Kurawae E.B., Laigret P., Lambais M.R., Leite L.C.C.,  
 DR Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 DR Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 DR Marques M.V., Martins E.A.L., Martins E.M.P., Matsukuma A.Y.,  
 DR Meck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 DR Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 DR Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira X.A., Paris A.,  
 DR de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pereira J.S.,  
 DR Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.S.,  
 DR Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 DR de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 DR da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 DR da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza M.H.,  
 DR de Souza A.P., Terenzi M.F., Trujillo D., Tsai S.M., Ishiko M.H.,  
 DR da Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.J.,  
 DR Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 DR "The genome sequence of the plant pathogen Xylella fastidiosa";  
 DR Nature 406:151-159(2000).  
 RL NATURE 406:151-159(2000).  
 CC -I- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: AF003883; AAF83187.1; --  
 DR PIR: H82813; H82813.  
 DR HSP: O00441; 10XA.  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PRO0385; P450.  
 DR PROSITE: PS00086; CYTOCHROME P450; 1.  
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 KW Complete proteome.  
 FT METAL 348 348 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT SEQUENCE 402 AA; 45023 MW; 5891FE7EC128F63 CRC84;  
 Query Match 23.9%; Score 489.5; DB 1; Length 402;  
 Best Local Similarity 30.7%; Pred. No. 2.3e-24;  
 Matches 122; Conservative 70; Mismatches 167; Indels 39; Gaps 6;  
 QY 25 HPEPVYESMRKADPVSD--EENQVWGVFLYDDVKKVVGKELFSSCMPQQTSSIGNS----- 80  
 DB 14 NPFLYETLRAQAPFVSGISGNALMT--GRYSLVDSLLNRNMGKYMESMRVYGDSDADM 72  
 QY 81 -----IINMDPPKHTKIRSVNKAFTPRVWKQWEPRIQETDELIOKFQGRSEFDL 131  
 DB 73 PLFOAESRMFTINPPAHTHRLGLWQAFTGRESSEMRPLAIDTAHQIDNFEQKPSVDL 132  
 QY 132 VHDFSYPLPVIIVISEL-----LVPSAHEQFKAWSDLVSTPKDSBEAEKAF 181  
 DB 133 VAEFAFPFPMQIICKMMDVIGDVAITGLAVSKIAK-----VFDSPEMSADE-----L 180  
 QY 182 EERDKCEELAAFPAGIIEKRNKPEQDIISILVEABETGEKLSGELIPCTLLLVAGN 241  
 DB 181 VHASTRYEELAQYFTKLIELRTHPGTDLISMLFRAEEDGKLTDEIVSNVIMGLIAGY 240  
 QY 242 ETTTNLSINAMYSILETPGVTEELRSHPELMPQAVEALRFRAPVLRRAKRDTEIGG 301  
 DB 241 ETTTNNMIGNALIALHRHPEQLALKSLDSIMPQAVSECLRYDGSVQVFTYRAAMDDIEVBG 300  
 QY 302 HLIKEGDMVLAFVASANDEAKFDRPHMFDIRRHNPPIAFGHGIFHCLGAPLARLEANI 361  
 DB 301 ELVPRGTVVFLMGLSANKNDPAQFTHPDQDLITRKQRLQSGFAGIHHCGLYRLAJELEEC 360  
 QY 362 ALTSLISAFHMECVSITP-----ENSIVIVGLKSFV 396  
 DB 361 AUTTLFERPLHLALHLDALNWNQSNLGVNTLIVLD 398

Wed Jun 2 13:01:00 2004

Search completed: May 28, 2004, 23:18:05  
Job time : 42 secs

us-09-869-334b-1.rsp

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:12:54 ; Search time 3275 Seconds  
(without alignments)  
15762.306 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgttaaacgcgcg.....gcttcggtgaaatgttaa 1191

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1189.4	99.9	35739	1	AF015825 Bacillus
2	1189.4	99.9	201241	1	BSUB00007 Z99110 Bacillus su
3	425	35.7	966	6	AX433775 Sequence
4	242	20.3	1213	6	AX433767 Sequence
5	234.4	19.7	290685	1	AE017032 Bacillus
6	229.8	19.3	3187	1	M24523 B.subtilis
7	182.6	15.3	4317	1	BACRTP Z21972 B.megateriu
8	169	14.2	1902	1	BMCTP450A X16610 Bacillus me
9	165.8	13.9	6629	1	AB072568 Streptomy
10	162.4	13.6	5608	1	BSCYPAZL Y11043 Bacillus su
11	162.4	13.6	17828	1	BSU93876 U93876 Bacillus su
12	162.4	13.6	199173	1	BSUB0014 Z99117 Bacillus su
13	159.6	13.4	300893	1	AE017006 AE017006 Bacillus
14	156.4	13.1	10167	1	SFU08223 U08223 Streptomyce
15	150	12.6	342300	1	MLEPRTN8 AJ583924 Mycobacte
16	149	12.5	1233	6	AX697983 Sequence
17	149	12.5	60196	6	AX697977 Sequence
18	145.8	12.2	17512	1	AB071405 Lechevali
19	145.8	12.2	25681	1	SAE414559 Saccharot
20	145.8	12.2	26144	1	AB090952 Lechevall
21	145.8	12.2	28654	1	AF534707 Lechevall
22	144.2	12.1	3078	1	SRCPC450A M83110 Saccharopol
23	144	12.1	39228	1	MLCB1788 AL008609 Mycobacte
24	139.2	11.7	8478	6	AR173058 Sequence
25	139.2	11.7	10153	1	BSU51868 US1868 Bacillus su
26	139.2	11.7	199922	1	BSUB0016 Z99119 Bacillus su
27	139.2	11.7	220060	1	AF008220 Bacillus
28	138.2	11.6	207829	1	BSUB0010 Z99113 Bacillus su
29	136.4	11.5	6739	1	AB088066 Bacillus
30	135.4	11.4	293175	1	AP005023 Streptomy
31	125	10.5	300350	1	AF006574 Gloebacte
32	124.2	10.4	145911	1	AP003014 Mesorhizo
33	122.4	10.3	309050	1	SC0939117 AE011961 Xanthomon
34	120.6	10.1	10417	1	AE011961 AE012559 Xylella f
35	120.6	10.1	300885	1	AE012559 Agrobacte
36	119.2	10.0	10029	1	AE009087 Agrobacte
37	119.2	10.0	10184	1	AE008053 Agrobacte
38	118.4	9.9	154620	2	BX323047 Danio rer
39	118.2	9.9	16057	1	AE007165 Mycobacte
40	118.2	9.9	349606	15	BX842583 Mycobacte
41	117.8	9.9	24268	1	AB070947 Streptomy
42	117.8	9.9	308550	1	AP005030 Streptomy
43	117	9.8	8407	1	SEU082823 U82823 Saccharopol
44	116.4	9.8	18109	1	AE006971 Mycobacte
45	116.4	9.8	320050	1	BX248336 Mycobacte

# ALIGNMENTS

RESULT 1  
LOCUS AF015825 35739 bp DNA linear BCT 07-MAY-1998  
DEFINITION Bacillus subtilis 168 cot1-rapA region sequence.  
ACCESSION AF015825  
VERSION AF015825.1 GI:2612880  
KEYWORDS  
SOURCE Bacillus subtilis  
ORGANISM Bacillus subtilis  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
REFERENCE 1 (bases 1 to 35739)  
AUTHORS Rivolta,C., Soldo,B., Lazarevic,V., Joris,B., Maue,C. and Karamata,D.  
TITLE A 35,7 kb DNA fragment from the Bacillus subtilis chromosome containing a putative 12.3 kb operon involved in hexuronate

catabolism and a perfectly symmetrical hypothetical  
catabolite-responsive element

JOURNAL  
MEDLINE

9573062

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 35739  
/organism="Bacillus subtilis"  
/mol\_type="genomic DNA"  
/strain="168"  
/db\_xref="taxon:1423"  
/map="cot-rapA region"  
complement (1..129)  
/gene="cotJ"  
complement (1..114)  
/gene="cotJ"  
/codon\_start=1  
/transl\_table=1  
/product="spore coat protein"  
/protein\_id="AAC46305.1"  
/db\_xref="GI:2612881"  
/translation="MDYPLNEQSFEQITPYDERQYYPYRPRPPFPYYPY"  
complement (122..129)  
/gene="cotJ"  
355..1806  
/gene="yJfA"  
355..361  
/gene="yJfA"  
368..1771  
/gene="yJfA"  
/note="similar to Rhizobium leguminosarum NodB protein"

RBS

gene

RBS

CDS

/codon\_start=1  
/transl\_table=1  
/product="NodB-like protein"  
/protein\_id="AAC46306.1"  
/db\_xref="GI:2612882"  
/translation="MLAKRIKWHVLIADVGVVGLGIFHNHSLKKEIVMKNKVRTPDQ  
YGNVETAVNCKTNYAVVYFVKRMDKSKFASKEVQFQKQKTDVQDHTT  
KNELVYDKIVHYAKQTVAVFNEVYKLGAGHQTVKTFNDFSKOAFSLIDDIFK  
EADYILKSLIAYHSLKKNKDIADALLKEGTAPKENSFEALKEDYIELYDFTY  
QVAGYLGQSIAKSLKDLKQIIDKAKNKIKKQPKHEVISLPEKTVDPN  
QVIALTFDDGNPATNQILSLKKYKGHATFFVIGSRVQYYPETLIRLMKEGNEVG  
NHSWHPILTRLSVKALKQINDTDIIEKISGRPTLVPPYGGINDELRSQMDV  
ALWVDVDPDKORNKKTIVDRVVMNQAGDGRITLIDHIRTSDADADEIINKLTDQGYQ  
LVTVSQLEEVKKQREAK"  
1775..1806  
/gene="yJfA"  
complement (1811..2298)  
/gene="yJfA"  
complement (1811..2284)  
/gene="yJfA"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAC46307.1"  
/db\_xref="GI:2612883"  
/translation="MKRLPMKASLVLFAVVFVAVKGAAPAKAETHAYDGKSPYNDCA  
SSGSTKSSNLVNASQVIGVELKFSSTCKTAWAKITMNTLTSGYEANAEITRND  
GKYNCDGAGNGKAVAGQKSCYTPWVYLDLPRTSFAFKYSGPNLNVWATTSY"  
complement (2291..2298)  
/gene="yJfA"  
complement (2363..2589)  
/gene="yJfB"  
complement (2363..2399)  
/gene="yJfB"  
complement (2409..2576)  
/gene="yJfB"

stem\_loop

gene

CDS

RBS

gene

stem\_loop

CDS

/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAC46308.1"  
/db\_xref="GI:2612884"  
/translation="MDIPALSVMHQASLAQNVIATLTKRLDTAQNADOTLKMIOH  
PTLGQTIDVKA"  
complement (2581..2589)  
/gene="yJfB"  
2689..3632  
/gene="yJfC"  
2689..2696  
/gene="yJfC"  
2703..3632  
/gene="yJfC"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAC46309.1"  
/db\_xref="GI:2612885"  
/translation="NNLVMEKTFEQYEKLFSEEQKREDVRYTMMRPFKMTAIQV  
PLKAKPNQYDVIMAAKMLGYLDVDAESQKALQILKESHVSETAESALRQCIFAE  
RQLRVNAKEIKFGLVADPNKLQKQXGFGGIQFTHVWPNNSYMLPRIPSIIA  
HEFHNVRSYIDFHGGSVSGDYLYIEGLAESFARELCTEELGPMVTFDEHLYQY  
SIDVINEVLIDVGFSEVSMFGDPIANDQGFAPGLSAPAGYAVGHAVQSFNQCCH  
ITISEATRLDAKTIISQCGLFLHEKSVTVSGAF"  
3602..3630  
/gene="yJfC"  
complement (3610..4024)  
/gene="yJfA"  
complement (3610..4008)  
/gene="yJfA"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAC46310.1"  
/db\_xref="GI:2612886"  
/translation="MKRNMIYAVFTILIGLGRASFSSVLPDTLNTVGLSLSMA  
MFTGCGFLPRKLMTITIGTISLSCFCFVIEFSQLYHEAWIDQIRDTSLGSLVLYGFL  
WSDIEATIGIAACAALELLVLGKKRCM"  
complement (4020..4024)  
/gene="yJfA"  
complement (4084..4697)  
/gene="yJfB"  
complement (4084..4099)  
/gene="yJfB"  
complement (4109..4684)  
/gene="yJfB"  
/codon\_start=1  
/transl\_table=1  
/product="putative lipoprotein"  
/protein\_id="AAC46311.1"  
/db\_xref="GI:2612887"  
/translation="MKMTSAITAAAVTSCPTGFGNASFSAKAACTNLTSENIN  
QSAALVKNDINTAYKEMPPQQAQGLTINKSTGKDVHAAFGEPFVGGNREDLHW  
NMQPGYGSYHKDMTISEIRYFTGVRQLNIGGVTPVLQKQLGPNVLTVPFTD  
EIDYVDYGRYELHFVIGTIDQTAHVNLIKAK"  
complement (4690..4697)  
/gene="yJfB"  
4814..7787  
/gene="yJfC"  
4814..4822  
/gene="yJfC"  
4830..7787  
/gene="yJfC"  
/note="similar to formate dehydrogenase alpha subunit of  
Moorella thermoacetica; member of the 4Fe-4S ferredoxins  
family"  
/codon\_start=1  
/transl\_table=1  
/product="formate dehydrogenase alpha subunit homolog"  
/protein\_id="AAC46312.1"  
/db\_xref="GI:2612888"



```

Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
Galleron,N., Ghim,S.-Y., Glaser,P., Goffeau,A., Golightly,E.J.,
Grandi,G., Guiseppe,G., Guy,B.J., Haga,K., Haeche,J., Harwood,C.R.,
Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Huilo,M.F.,
Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
Klaer-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
Koningsbein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
Lardinois,S., Lauber,J.S., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
Masuda,S., Mauei,C., Medigue,C., Medina,N., Mellado,R.P.,
Mizuno,M., Moesti,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
Ogawa,K., Ogiwara,A., Cudega,B., Park,S.H., Parro,V., Pohl,T.M.,
Portetlelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadale,Y., Sato,T.,
Scanlan,E., Scheich,S., Schroeter,R., Scoffone,P., Sekiguchi,J.,
Sekowska,A., Seteri,S.J., Serror,P., Shin,B.S., Solido,B.,
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
Takeuchi,M., Tanakoshi,A., Tanaka,T., Terptria,P., Tognoni,A.,
Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
Viari,A., Wambat,R., Wedler,E., Wedler,H., Weitzeneget,T.,
Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
Danchin,A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377
2 (bases 1 to 201241)
Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
Direct Submission
Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
48 89 48
On Jul 7, 2003 this sequence version replaced gi:2633472.
This entry contains data from release R16.1 of the Subtilist
database. Further data on gene annotation and detailed information
about changes from previous releases can be found at
http://genolist.pasteur.fr/Subtilist/.
Location/Qualifiers
1..201241
/organism="Bacillus subtilis subsp. subtilis str. 168"
/mol_type="genomic DNA"
/strain="168"
/db_xref="taxon:224308"
67..852
/_gene="yjaZ"
/_locus_tag="BSU11350"
67..852
/_gene="yjaZ"
/_locus_tag="BSU11350"
/_function="unknown"
/_note="similar to unknown proteins"
/_codon_start=1
/_transl_table=11
/_protein_id="C26314992.1"
/_db_xref="GI:26334489"
/_db_xref="SPTREMBL:O31596"
/_translations="MSVEQYFSLWRAGSIDDLAHYIILPLFGSAEKNKWKGLIHGHQ
HGMFKNKEGIOTVSKLKGFGYHKGKQVGNKMGQDPVIVTLFVDERNRRLRII
EPFSGKSLAQDQKMFLLSSDLDPGVSALMTWHEHYHVCRLGHILTKEDNVTLLDTII
MEGLAEYATVFERGRSTQSWTTWTFPEQQLAYEKKIAPHRIIDKRDNLFPQLLGGK
GYQPKMLGYAVGVNIVKKYILTASKASTADGSLSPAEITFLDAML"
1044..2030
/_gene="appB"
/_locus_tag="BSU11360"
1044..2030
/_gene="appB"
/_locus_tag="BSU11360"
/_function="oligopeptide transport"
/_codon_start=1

```



gene	/translation="MAAYIIRRTLMSPILGILGITILTSFVIMKAAAPGDPMTLAWDPKILS QADREQLEYGLNDPQYQYLKWNMGWDGFGTSIVRKGTPVSLIWARLPNTLLIL MLVSTIILAMISIPFGVLSAKPYSKIDYGITFTSFIGLAIPNFEGLLIILVLSVNL WRPPTGGVLTLETFENIFDRIHLLILPAFLATADMAGLTRVTRSNMLDVLINQDYLR GAQFKENRVLPEKGLERNALLPVITIFGLMIPSFITGGVSVVEQIFTPWPGKGLFYDS AFQKTIIRIMAMTVISAVLVVGNLIADLYIAVDPRIEY" 5781..6692 /gene="appC" /locus_tag="BSU11400" 5781..6692 /gene="appC" /locus_tag="BSU11400" /function="oligopeptide transport" /codon_start=1 /transl_table=11 /product="oligopeptide ABC transporter (permease)" /protein_id="CAB12997.1" /db_xref="GI:2633494" /db_xref="GOA:P42063" /db_xref="SWISS-PROT:P42063" /translation="MSELOTTPSEIRLKENISKKPTMTKIPWEKSKNNKLAILGAV ILFIIMSAVPAPLIAVPOQQLDKYKAPGLEHLMGTGKGRDIPSFILYGARVS LLVGFASVGSILGTVTARLGVGFGRVDIVMRVVDIVLSIPDIPILLITLVTFKP GVDKLLILFCLTGHVTVARLVGRGFSILRSREYVLAAKTIGTTHKILFSHILLNALG PIIVSATLKVGSVIIAESAISYLGFQPTPIASGNWMLQDAQNFTWIQAWVPELPFG LFIIMTVLCFNFVGGLNDLADPKNIK" 6898..7650 /gene="yJbA" /locus_tag="BSU11410" 6898..7650 /gene="yJbA" /locus_tag="BSU11410" /function="unknown" /note="similar to unknown proteins" /codon_start=1 /transl_table=11 /protein_id="CAB12998.1"	
CDS		
gene	Query Match	99.9%; Score 1189.4; DB 1; Length 201241;
CDS	Best Local Similarity	99.9%; Pred. No. 0;
	Matches 1190; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1	ATGAATGTGTTAAACCGCGGCHAGCTTGACGAGCGCTGCTCAATGGGAAACAAA 60
DB	80913	ATCAATGTGTTAAACCGCGGCAAGCTTTCGACGAGCGCTCTCAATGGGAAACAAA 80972
QY	61	CAGGATGCGTATCATCCGTTTCCATGTATGAATCATGAGAAAGATGCGCTGTTCC 120
DB	80973	CAGGATGCGTATCATCCGTTTCCATGTATGAATCATGAGAAAGATGCGCTGTTCC 81032
QY	121	TTTGATGAAGAAAAACCAAGTGTGGAGCGTTTTCTTTATGATGATGTCAAAAAAGTTGTT 180
DB	81033	TTTGATGAAGAAAAACCAAGTGTGGAGCGTTTTCTTTATGATGATGTCAAAAAAGTTGTT 81092
QY	181	GGGGATAAGAGTTGTTTTCCAGTTTCATGCGCGACGACAGCTCTATTGGAAATCC 240
DB	81093	GGGGATAAGAGTTGTTTTCCAGTTTCATGCGCGACGACAGCTCTATTGGAAATCC 81152
QY	241	ATCATTAACATGACCGCGCGAAGCATACAAAATCCGTTTCAGTGTGTGAACAAAGCCTTT 300
DB	81153	ATCATTAACATGACCGCGCGAAGCATACAAAATCCGTTTCAGTGTGTGAACAAAGCCTTT 81212
QY	301	ACTCOCGCGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATACAGATGAATCTGATT 360
DB	81213	ACTCOCGCGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATACAGATGAATCTGATT 81272
QY	361	CAAAAAATTTACGGGGCGAGTGATTGACCTTGTTTCAGATTTTTCATACCCGCTTCG 420
DB	81273	CAAAAAATTTACGGGGCGAGTGATTGACCTTGTTTCAGATTTTTCATACCCGCTTCG 81332
QY	421	GTTATTGTGATATCTGAGCTGCTGGGAGTGCCCTTCAGCGCAGATGCAACAGTTTAAAGCA 480
DB	81333	GTTATTGTGATATCTGAGCTGCTGGGAGTGCCCTTCAGCGCAGATGCAACAGTTTAAAGCA 81392

QY	481	TGGTCTGATCTTCCTGGTCAGTACACCGAAGATAAAAGTGAAGAAGCTCAAAAAAGCCTTT	540
Dd	81393	TGGTCTGATCTTCCTGGTCAGTACACCGAAGATAAAAGTGAAGAAGCTCAAAAAAGCCTTT	81452
QY	541	TTGGAAGAACGAGATAAGTGTGAGAGAAACTGGCCGCCTTTTTTGCCCGCATCATAGAA	600
Dd	81453	TTGGAAGAACGAGATAAGTGTGAGAGAAACTGGCCGCCTTTTTTGCCCGCATCATAGAA	81512
QY	601	GAAAAGCGAAACAACCCGACAGGATATTATTTCTATTTTAGTGAAGCGGAAGAAACA	660
Dd	81513	GAAAAGCGAAACAACCCGACAGGATATTATTTCTATTTTAGTGAAGCGGAAGAAACA	81572
QY	661	GGCGAAGAGCTGTCGGGTGAACAGCTGATTTCGGTTTGTCAAGCTGCTGCTGTGGGCCGA	720
Dd	81573	GGCGAAGAGCTGTCGGGTGAACAGCTGATTTCGGTTTGTCAAGCTGCTGCTGTGGGCCGA	81632
QY	721	AATGAAACCCTACAAAACCTGATTTCAAATGGATGTACAGCANATTTAGNAAGCCGAGGC	780
Dd	81633	AATGAAACCCTACAAAACCTGATTTCAAATGGATGTACAGCANATTTAGNAAGCCGAGGC	81692
QY	781	GTITACGAGGAAGTGGCGAGCCATCTGTAACATGATGCTCAGCGAGTGGAGGAAGCCTTG	840
Dd	81693	GTITACGAGGAAGTGGCGAGCCATCTGTAACATGATGCTCAGCGAGTGGAGGAAGCCTTG	81752
QY	841	CGTTTCAGAGCGCCGCCGCCGTTTGGAGCGCATTTGCCAAGCGGATACGAGATCGGG	900
Dd	81753	CGTTTCAGAGCGCCGCCGCCGTTTGGAGCGCATTTGCCAAGCGGATACGAGATCGGG	81812
QY	901	GGGCGACCTGATTTAAAGAAGGTGATATGGTTTGGCGTTTCTGGCATCGGCAATCGTGAT	960
Dd	81813	GGGCGACCTGATTTAAAGAAGGTGATATGGTTTGGCGTTTCTGGCATCGGCAATCGTGAT	81872
QY	961	GAAGCAAGTTTGAACAGCGCACATGTTGATATCGCCGCCCATCCCAATCCGCATATT	1020
Dd	81873	GAAGCAAGTTTGAACAGCGCACATGTTGATATCGCCGCCCATCCCAATCCGCATATT	81932
QY	1021	GGTTTGGCCACCGGCATCCATTTTSCCTTGGGCGCCCGCTTGCCCGCTTTGAAGCAAAAT	1080
Dd	81933	GGTTTGGCCACCGGCATCCATTTTSCCTTGGGCGCCCGCTTGCCCGCTTTGAAGCAAAAT	81992
QY	1081	ATCGCGZTTAAACGCTTTTGATTTCTGCTTTCCCTCATATGAGTCGCTCAGTATCACTCCG	1140
Dd	81993	ATCGCGZTTAAACGCTTTTGATTTCTGCTTTCCCTCATATGAGTCGCTCAGTATCACTCCG	82052
QY	1141	ATTGAAAACAGTGTGATATACGATAAAGAGCTTCCGCTGTGAAAATGTAA	1191
Dd	82053	ATTGAAAACAGTGTGATATACGATAAAGAGCTTCCGCTGTGAAAATGTAA	82103
 RESULT 3 AX433775 LOCUS 966 bp DNA linear PAT 28-JUN-2002			
Sequence 2190 from Patent WO0229113.			
ACCESSION AX433775			
VERSION AX433775.1 GI:21658583			
KEYWORDS Bacillus licheniformis			
SOURCE Bacillus licheniformis			
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
REFERENCE 1 Berka,R. and Clausen,I.G.			
AUTHORS Methods for monitoring multiple gene expression			
TITLE Patent: WO 0229113-A 2190 11-APR-2002;			
JOURNAL Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)			
FEATURES Location/Qualifiers			
source 1..966			
/organism="Bacillus licheniformis"			
/mol_type="unassigned DNA"			
/db_xref="taxon:1402"			
ORIGIN			
Query Match 35.7%; Score 425; DB 6; Length 966;			



QY	855	GGCCCGCGTTTGGAGCGCATGCGCAAGCGGGGATACGAGATCGGGGGGCACCTGATTAA	914
DB	866	CTGCAAGCGATCGCGCGACCGCGCGAGAGAGCTTGATATCGGAGCGGTGAGGATTGC	925
QY	915	AGAAAGTGATATGTTTGGCGTTTGGCGATCGGCAATCGTGATGAACAAAGTTTGA	974
DB	926	MAAGGTTCTACAGTGATCAGCTGGTGGTTCGCTTCAGCGAATCGTGACAACTTAAGTTTGA	985
QY	975	CAGACCGGACATGTTTGATATCGCGCCCATCCCAATCGCATATTCGTTTGGCCACCG	1034
DB	986	CGATCTGACAGCTTCAAGCTTGATCGCAATCAACCTCATATGAGCTTCGGCTCGG	1045
QY	1035	CATCCATTTTGGCGGCCCGCTTCCCGCTTGAAGCAAAATATCGGTTAAGCTC	1094
DB	1046	CATCCATTTTGGCTCGCGCTCCCTCGCGGCTTGAAGCAAAAGTCCGCTCGATTAA	1105
QY	1095	TTTGAT 1100	
DB	1106	CTTGCT 1111	
RESULT 5			
LOCUS			
DEFINITION	AE017032	290685 bp DNA linear	BCT 30-APR-2003
ACCESSION	AE017032		
VERSION	AE017032.1	GI:30257133	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			



QY 988 TTGTATATCCGCGCATCCCAAT---CGCATATTCGCTTTGGCCACGGCATCCATTTT 1044  
 DB 124502 TTTATATACATCGATAGGAATGAAGACATTTAACTTTGGTAAAGGCCCTCACITT 124561  
 QY 1045 TGCTTTGGGCGCGCTTCCCGCTTCTGAAGCAATATCGGTTAACTGTTTGAATTTCT 1104  
 DB 124562 TGCTTTAGGCGCACCACTTGGCGGTTTGAAGCTGAGATTGCATTAACCTTTATAAT 124621  
 QY 1105 GCTTTTCTCATATGG 1120  
 DB 124622 GCTTTTGAAGAAGATAG 124637

RESULT 6  
 BACKUP/C 3187 bp DNA linear BCT 26-APR-1993  
 LOCUS B.subtilis rtp gene, complete cds and proc gene (put.), 5' end.  
 DEFINITION M24523 M36988  
 ACCESSION M24523.1 GI:143477  
 VERSION DNA-binding protein; proc gene; replication terminator protein;  
 KEYWORDS terC gene.  
 SOURCE Bacillus subtilis  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1 (bases 1 to 3187)  
 Lewis, P.J. and Wake, R.G.  
 DNA and protein sequence conservation at the replication terminus  
 in Bacillus subtilis 168 and W23  
 J. Bacteriol. 171 (3), 1402-1408 (1989)  
 MEDLINE 89155440  
 PUBMED 2493444  
 REFERENCE 2 (sites)  
 Ahn, K. and Wake, R.G.  
 A unique open reading frame adjacent to the replication terminus of  
 the Bacillus subtilis W23 chromosome compared with Bacillus  
 subtilis 168  
 Unpublished (1990)  
 REFERENCE 3 (sites)  
 Ahn, K.S. and Wake, R.G.  
 Variations and coding features of the sequence spanning the  
 replication terminus of Bacillus subtilis 168 and W23 chromosomes  
 Gene 98 (1), 107-112 (1991)  
 MEDLINE 91192601  
 PUBMED 1849493  
 COMMENT Original source text: B.subtilis (strain W23) DNA.  
 Draft entry and computer-readable sequence for [2] kindly submitted  
 by R.G.Wake, 24-JUL-1990.  
 Author address: R.G.Wake  
 University of Sydney, Dept. of Biochemistry  
 NSW 2006, AUSTRALIA  
 E-mail: GERRY@EXTRO.UCC.SU.OZ.AU.  
 Location/Qualifiers  
 1. .3187  
 /organism="Bacillus subtilis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1423"  
 1. .1499  
 /organism="Bacillus subtilis"  
 /mol\_type="genomic DNA"  
 /strain="W23"  
 /db\_xref="taxon:1423"  
 complement (120. .1337)  
 /note="ORF 405"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAA22720.1"  
 /db\_xref="GI:143478"

FEATURES  
 source  
 source  
 CDS  
 1461..1615  
 /note="inverted repeat"  
 1642..2010  
 /note="replication terminator protein (rtp; terC)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAA22721.1"  
 /db\_xref="GI:143479"  
 /translation="MKERKRSSTGFLVKQKAPFKIKMINTTEERLYGLKLLVLRSE  
 FKIEGKPNHTEVYRSLHLLDGLKQIKVKEGAKLQEVVLYQFYAEAKLYKQ  
 LKVELDRCKKLIKALSDNF"  
 2156..2929  
 /note="ORF 257 proc protein (put.); putative"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAA22722.1"  
 /db\_xref="GI:143480"  
 /translation="MRKTKTKMLPIFDKQKAFIAGSMAEGMISGIVRANKIPKQ  
 NICVTRNTERLAELEQYIKGASPNQICIEDMDVLILAMPKDAESALSSLKTRI  
 OPHOLILSVLAGITTFIEQSLLNQPVVRVMENTSMWISGASATIALGKVKVSEDLQ  
 LAFALLGCMGEVTTIOENQMDIFGTIAGSPAFYILMEFIEKTEGEAGLDKLSRSI  
 GAOTLLGAARKMLMETGEQPEVLRDNITSPNGTTAAGLQA"

# ORIGIN

Query Match 19.3%; Score 229.8; DB 1; Length 3187;  
 Best Local Similarity 53.8%; Pred. No. 4.9e-55;  
 Matches 474; Conservative 0; Mismatches 407; Indels 0; Gaps 0;  
 QY 241 ATCATTAACATGACCCCGCAACATACAAATATCCGTTTCAGTCGTGAACAAAGCCGTTT 300  
 DB 1091 ATATTAACAATGACCTCCGAACACACCAATATACGGGCTCTCGTCAGCAGACGCTT 1032  
 QY 301 ACTCCGCGCGTATGAAGCAATGGAAACCGAGAAATTCAGAAATTCACAGATGAATGAT 360  
 DB 1031 ACCCCAAAAGCAGTAAACAACTCGAAACCCGCATAAAGACGTGACGCGTTTCTCTTA 972  
 QY 361 CAATAATTTAGGGGCGCAGTGGTTTACCTTTTTCACGATTTTTCATACCGCTTCG 420  
 DB 971 CAAGAAGCAGCGCAAAAAGACACCAATATCATATGAAGATTTTCAGAGTCCCTCCCT 912  
 QY 421 GTTATTGTATATCTGAGCTCTGGAGTGCCTTCAGCGCAGATGGAACAGCTTTAAAGCA 480  
 DB 911 GTTATCATATAGCTGAATGCTGGCGCTCGATTGAAGACCGCCACTTGAATTAACA 852  
 QY 481 TGGTCTGATCTTCTGTCAGTACACCGAAGATTAAGTGAAGAGCTGAAAGACGCTTT 540  
 DB 851 TACTCCGACGTACTTGTAGCTGGAGCAAGGACAGCTCCGCAAAAGCTGTTGCTGACATG 792  
 QY 541 TTGGAAGACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTCGCGGATCATAGAA 600  
 DB 791 GTCCATAATCGGCGTGACGCGCCACGCTTTTAACTGACTATTCAGAGACATCTGTCA 732  
 QY 601 GAAGAAGCAAAACAAACCGGAACAGGATATTTCTATTTTAGTGGAAAGCGGAAGAAACA 660  
 DB 731 AAACGAGAGCTGAACCAACCAAGAGATTTAATGACATGCTTTTACAGCGCAATTCGAT 672  
 QY 661 GCGGAGAGCTGTCCGTTGAAGAGCTGATTCGTTTTCACGCTGCTGCTGCTGCGCGGA 720  
 DB 671 GGAGAAATATTTAAACAGAAAGACAGTGTATCGGTTTTTTTGTATTTCTTCTCTCGT 612  
 QY 721 ATGAAACCACTACAAACCTGATTTTCAATGCGATGTACAGCATATTAAGAAACGCGAGC 780  
 DB 611 AACGAAACGACTACAAATTTGATCGCAATGAGTTCGCTATCTGACAGAAATTCAGTG 552  
 QY 781 GTTTACGAGGAACTCCGACGCACTCTGAACTGATGCTTCAGCGAGTGGAGAAAGCTTG 840  
 DB 551 GTTCAACAGCAGGTGAACAAATACAGCAATGTTGCCAATGTAATGAAGAAACGCTT 492  
 QY 841 CGTTTCAGAGCGCGCGCGCGCTTTTGGAGCGCATTCGAGCGGATACCGGATACCGG 900  
 DB 491 CGCTACTATTCGCTGTGAGGCGCATTTGGCGCTGTGGCGAGGACACAGAGCTTGGC 432

```
QY 901 GGGCACCTGATTAAAGAGTGATATGTTTGGCGTTTGTGCGATCGCGAAATCGTGAT 960
D 431 GGAGTGTTTCATCAAAAGAGGATCTTCOGTTATCAGCTGGATCGGTCAGCCATCGGAT 372
QY 961 GAAGCAAGTTTGACAGACCGACATGTTTGATATCGCGCCATCCCAATCGCATAT 1020
D 371 GAAGTAATTCGTGAACCTGATGTTTATAGATTGACCGCCCTCTCTCATCTC 312
QY 1021 GCGTTTGGCCACCGCATCCATTTTCCCTTGGGCGCCCGCTTCCCGCTTGTGAGCAAT 1080
D 311 AGCTTGGATTCCGGATTCACTTTTCCCTCGCGCTCCATTCGCGAGGCTAGAAGCAAC 252
QY 1081 ATCGGTTAAAGCTTGTGATTTCTGCTTCTTCATATGA 1121
D 251 ATCGCTCTTTCTCCCTCTTATCAATGTCAGCTTGATCGA 211

RESULT 7
LOCUS BMCTP450A 4317 bp DNA linear BCT 13-OCT-1993
DEFINITION B.megaterium cytochrome P450meg, ORF1 and ORF2 genes.
ACCESSION Z21972
VERSION Z21972.1 GI:288298
KEYWORDS cytochrome P450meg; ORF1; ORF2.
SOURCE Bacillus megaterium
ORGANISM Rauschenbach, R., Isernhagen, M., Noeske-Jungblut, C., Boidol, W. and
REFERENCE 1 (bases 1701 to 3400)
AUTHORS Siewert, G.
TITLE Cloning sequencing and expression of the gene for cytochrome
P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium
ATCC 13368
JOURNAL Mol. Gen. Genet. 241 (1-2), 170-176 (1993)
MEDLINE 94049677
PUBMED 8232201
REFERENCE 2 (bases 1 to 4317)
AUTHORS Siewert, G.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1993) Siewert G., Schering AG, Muellerstr.
170-178, D-1000 Berlin 65, Germany
FEATURES
source 1..4317
/organism="Bacillus megaterium"
/mol_type="genomic DNA"
/strain="ATCC 13368"
/db_xref="taxon:1404"
37..717
/gene="ORF1"
37..717
/gene="ORF1"
/citation=[1]
/codon_start=1
/transl_table=11
/db_xref="CAA79984.1"
/db_xref="GI:288298"
/db_xref="SPTREMBL:Q06073"
/translation="MRISLKKKALQSLGGKXGTVTVSLMLLLFLNLVPLIVEVIGS
GGFSEWLMQBEETPLWSDIFSVMVSIALLPLTISTWFLNLVREGNPGIPEVFAIKD
KTSFKLIGASILQAIIFLMSLLLIIFGIKAIAYSQQFLLKDPHTVLEATES
KRMKGLKWKVFLMHLUSFIGWILGILGLLWLIIPVAGTTTAAFYBELLVQEDI
DDQQLIEG"
724..748
/citation=[1]
/repeat_region /rpt_type=INVERTED
/rpt_unit="724..733"
751..794
/citation=[1]
/repeat_region /rpt_type=INVERTED
/rpt_unit="751..771"
1330..1373
/citation=[1]
/repeat_region

/rpt_type=INVERTED
/rpt_unit="1330..1347"
1976..2001
/citation=[1]
/rpt_type=TANDEM
/rpt_unit="1976..1988"
2008..2013
/citation=[1]
2024..3256
/function="steroid-15beta-monooxygenase"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="cytochrome P450meg"
/protein_id="CAA79985.1"
/db_xref="GI:288300"
/db_xref="GOA:Q06069"
/db_xref="SWISS-PROT:Q06069"
/translation="MKEVIAVKEITRFKTRTEBFSYAMCKRMLENDPVSYHEGTDTW
NFKYBDVKVLSYDKHFSVRKRTTISVGTDSSEGSVPKIQITSDSPDHRKRSL
LAAAFTRSLQNWEPRIQETADELIGQMDGTEIDIVASLASPLIIVMADLMGVPSK
DRLLFKKVDVTLFPDREKQEVDDKLAQAAKEYQYLYPIVYQKELNADDIISDL
LKSEVDGEMFTDDDEVVVTWMLIIGAGVETSHLLANSFYSLXDDKKEVYELHNLDL
VPOAVEMLRFRFNLIKLDRTVKEDNDLLGVLEKSGSVVVMASAAANDDEMFEDPT
LNTHRPNNKHLTFGNGPHEFCIGAPLARLEAKIALTAFLAKKFKHIEAVSFQLEENLT
DSATGQTLTSLPLKASRM"
3256..3303
/citation=[1]
/function="terminator"
/rpt_type=INVERTED
/rpt_unit="3256..3275"
3560..4210
/gene="ORF2"
3560..4210
/gene="ORF2"
/citation=[1]
/codon_start=1
/transl_table=11
/protein_id="CAA79986.1"
/db_xref="GI:288301"
/db_xref="SPTREMBL:Q06074"
/translation="MKLQQLTIATFIISLVLSLIGFSMAFTISANEYLKFDVIVSLV
QWESPLLTDMKEFTYIGTASPIILSLVILFYLRLKHLRLVLFITAVMGSELL
NLWKLFPQARPDHLRLIDIGGYSFSGHAMNAYSILGILTFLMRHITARWARILL
ILFSWMLLSIGISRIYLVGHPYSDIILAGLGGCNIAISLWFFQRYQDRKKNDR"

repeat_region
/citation=[1]
/function="terminator"
/rpt_type=INVERTED
/rpt_unit="3256..3275"
3560..4210
/gene="ORF2"
3560..4210
/gene="ORF2"
/citation=[1]
/codon_start=1
/transl_table=11
/protein_id="CAA79986.1"
/db_xref="GI:288301"
/db_xref="SPTREMBL:Q06074"
/translation="MKLQQLTIATFIISLVLSLIGFSMAFTISANEYLKFDVIVSLV
QWESPLLTDMKEFTYIGTASPIILSLVILFYLRLKHLRLVLFITAVMGSELL
NLWKLFPQARPDHLRLIDIGGYSFSGHAMNAYSILGILTFLMRHITARWARILL
ILFSWMLLSIGISRIYLVGHPYSDIILAGLGGCNIAISLWFFQRYQDRKKNDR"

Query Match 15.3%; Score 182.6; DB 1; Length 4317;
Best Local Similarity 51.1%; Pred. No. 2.7e-41;
Matches 567; Conservative 0; Mismatches 494; Indels 48; Gaps 4;

QY 52 AAAACAAACAGGATCGGTATCATCCGTTTCATGATGATGATCGATGAGAGGATCG 111
D 2063 AAAACAAAGACGGAGGAATTTAGCCCGTACCGTTGTTAAAGGATGTTAGAAATGAC 2122
QY 112 CCGTTTCTCTTGATGACAAACCAAGTGTGAGCGTTTCTTTTATGATGATGTCMAA 171
D 2123 CCGTGTGAGTTATCACAGGAACCGATACGTGGAATGTCCTTAATATGAAGATGGAAG 2182
QY 172 AAAGTTGTGGGATAAAGAGTTGTTTTCAGTTCCATGCCCGCAGCAGCAAGCTTATT 231
D 2183 CGGGTTCTCAGTGATATAAACATTTTCAAGTGTTCGGAAACGGACGACGATTCAGTT 2242
QY 232 GGAAATTC-----ATCATTAACATGGACCG 258
D 2243 GGAACGATAGTGAGGAAGTTCTGTGCTCGAAAGATCCAAATCACTGAATCGGATCCA 2302
QY 259 CCGAGCATACAAAATCCGTTCCAGTCGTAACAAAGCCCTTACTCCGGCGCGTGATGAAG 318
D 2303 CCGTATCATAGAAACCGCGTTTCACTGCTGGCAGCAGCATTCACACTAGAGCTTCAA 2362
QY 319 CAAATGGGAACCGAGATTCAAGAAATCACAGATGAATCAAAATTTCAAAATTTAGGGCGCG 378
```





331 AGAATTCAGAAATCACAGATGAACCTGATTCACAAATTTCCAGGCGCAGTGCAGTTTAC 390  
|||||  
859 CGAATCAACAATTCAGAGACCTTGTAGAGCGATACAAAAGAAATCAACTATTAT 918  
|||||  
391 CTGTTACAGATTTTTCATACCGGCTTCGGTATTGTGATATCTGAGCTGCTGGAGTG 450  
|||||  
919 ATTGCGAGCATTTATCTTCTCTTTCTAGTCTAGTATAGCAGATCTATTTCGGCTGA 978  
|||||  
451 CCTTCAGCGCAGATGACAGATTTAAAGCATGCTGATCTTCTGCTCAGTACACCGAAG 510  
|||||  
979 CCGGTAAAGACCGATATCAGTTTAAATATGGGTAGACATCTCTT-----TCAG 1029  
|||||  
511 GATATAAGTGAAGAGCTGAAGAACCTTTTGGAGAACGAGATPAAGTGTGAGGAAGAA 570  
|||||  
1030 CTTATGACCAAGAAAGATTAGAGAGATTGAACAAGAAAGCAACGTCGAGGCTGAA 1089  
|||||  
571 CTGGCCGCTTTTTCGCGCATCATAGAGAAAGCGAAGCAACACCGGACAGATATT 630  
|||||  
1090 TATTTCATACCTCTACCCGATTTGCTATTGAAAGCGCTCCAACTTTCTGATGATATT 1149  
|||||  
531 ATTTCTATTATGAGGAGCGAAGCAACAGCGGAGAGCTGTCGGGTGAAGAGCTGATT 690  
|||||  
1150 ATCTCAGACTTAATTCAGCTGAAGTTGATGCTGAAACATTCACAGATGAAGAAATTGA 1209  
|||||  
691 CCGTTTTCACGCTCTCTGCTGCTGCGCGAATGAACCACTACAACTGATTTCAAAAT 750  
|||||  
1210 CACGCTACCATGTTGCTTCTAGTGCAGCGCTTGAACCCACAAGTCAAGCCATTGCAAT 1269  
|||||  
751 CGGATGACAGCATTTA---GAAACGCCGCGTTTACGAGGAAGTGGCAGGCACTCT 807  
|||||  
1270 ATGTTTATTTCTTTTATATGACCAAGTCTGTTATACGCGCTGAGAAACAATAGA 1329  
|||||  
808 GAATGATGCTCAGCGAGTGAAGAAAGCTTGCCTTTTCAGAGCGCCGCCCGTTTG 867  
|||||  
1330 GAATTAGCTCAAGAGAGTAGAAGAAATGCTTCATATCGTTTCATCTCCAGACGA 1389  
|||||  
868 AGGCGCATGCGAAGCGGATACGAGATCGGGGGGCGACCTGATTAAAGAGGTGATG 927  
|||||  
1390 GATCGACCGTTAAACAAGATAACGAGCTATTAGCGGTTTAACTAAAGAGGAGATGTC 1449  
|||||  
928 GTTTTGGGCTTTGGCATCGCAATCGTGTAGAGCAAGATTTGACAGCGGCACATG 987  
|||||  
1450 GTTATGCTGATGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1509  
|||||  
988 TTTGATATCCGCGCCATCCCAAT---CGCATATTGCTTTGGCCACCGCATTCATTT 1044  
|||||  
1510 GTAGACATTCATGCTTCAACTTAATAAAGCACTTAACATTCGAAACCGGACCTCATTT 1569  
|||||  
1045 TGCCTTGGGCGCCGCTTGGCGCTTGAAGCAATATCGCGTTAAACGCTTTGATTCT 1104  
|||||  
1570 TGTTTGGCGCCCTTTGCGCGACTTGAATGAAATTAATTTTGAAGCAATTTTAGAG 1629  
|||||  
1105 GCTTTTCTCATATGGA 1121  
|||||  
1630 GCTTTTCTCATATGGA 1646  
|||||

RESULT 9  
AB072568  
LOCUS  
DEFINITION  
Streptomyces virginiae visB, visB, visC, visD genes for L-lysine  
2-aminotransferase, 3-hydroxypicolinic acid:AMP ligase, lysine  
cyclodeaminase, cytochrome P450 monooxygenase, complete cds.  
AB072568.1 GI:18150120  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Streptomyces virginiae  
Streptomyces virginiae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1  
AUTHORS  
Namwat,W., Kamioka,Y., Kinoshita,H., Yamada,Y. and Nihira,T.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
source  
FEATURES  
Characterization of virginiamycin S biosynthetic genes from  
Streptomyces virginiae  
Gene 286 (2), 283-290 (2002)  
21940635  
11943483  
2 (bases 1 to 6629)  
Namwat,W., Nihira,T. and Kinoshita,H.  
Direct Submission  
Submitted (05-OCT-2001) Wises Namwat, Graduate School of  
Engineering, Osaka University, BioTechnology, Yamada-oka, Suita,  
Osaka 565-0871, Japan [E-mail: noi4bio.eng.osaka-u.ac.jp,  
Tel: 81-6-6879-7433, Fax: 81-6-6879-7432]  
Location/Qualifiers  
1..6629  
/organism="Streptomyces virginiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1961"  
complement (611..1870)  
/gene="visB"  
complement (611..1870)  
/gene="visA"  
/note="visA"  
/codon\_start=1  
/transl\_table=11  
/product="L-lysine 2-aminotransferase"  
/protein\_id="BAB83671.1"  
/db\_xref="GI:18150121"  
/translation="MTAAQRYLVPPFRRGSIQVQADLAALASVHSAEPLSGWNR  
FEQPREVTHALHSDVLTDTKAILLVHYGGIPADMAAILARARAGVWVE  
TDFCDVDPDTLMDSQITDTRKAILLVHYGGIPADMAAILARARAGVWVE  
DCAGMGASGSDGALGDIQCFSHSTKNTTLGEGMITLNADNARVERLRGN  
EYGVGFAPVGFADDEPALLPMWRFSADYVRQAVGISGAGTMTNMSAAAGVGLV  
DALERTERRRRAALDAASAFPRHPPHAGSGSTYHLYTFETVAGRGASDSC  
APLDRGVEVQLRYFPLHLLPEVAARPRSRGVRVQAEVWFDEHLNLPCHPLTDEQV  
DHLVESVTAAMHEATALTATGRLAAR"  
complement (1867..3522)  
/gene="visB"  
complement (1867..3522)  
/gene="visB"  
/note="visB"  
/codon\_start=1  
/transl\_table=11  
/product="3-hydroxypicolinic acid:AMP ligase"  
/protein\_id="BAB83672.1"  
/db\_xref="GI:18150122"  
/translation="MLDGCVPWPDAAERYEAGVWRGETLGLLRGWAERYGDRVAL  
VGDDGTRYRELDMMCDLAAAGPAGQVAPREVIVLPNTFSPVAVCFALFRLGAV  
PVFALPSVAVRSCGTWRSPVSRWIPDEVRGFDHRLVRLJLPELPRVAVFVAGD  
PGEGEDVLAALAEQDPVLPEDPADSDVAFLLSGITLALPKM:PRTHDDYAYQTRI  
TAGICELGEDTVYLAIVLVEFNPMGCPGILGVLGVRVVPAPDPTPOTCFPLIERE  
RYTMSVPTIVLHMLDAAGRTGHDLSLDVQVSSKLHDEVAARIEPELGVRLQOV  
FGMAGLITFTYDDPRQVLTTCQRPVSPADEIRVDDDDREVPDGTGELLTRGPY  
TLRGYRAPEHNASATADGFYRSGDVRRTPTGIVVGRVQVIRGDKVSAVEF  
EGHLLTAHPQVQAAVPAKPDVILGKICAYVVAADRGGAPTLPVRRLLRAGLADF  
LPDRLEVVEVPLTCLNKVKVLAARAGLLHSGGAGAGAR"  
3759..4646  
/gene="visC"  
3759..4646  
/gene="visC"  
/note="visC"  
/codon\_start=1  
/transl\_table=11  
/product="lysine cyclodeaminase"  
/protein\_id="BAB83673.1"  
/db\_xref="GI:18150123"  
/translation="METWLSERDTRVAVESVGRJALMDRVIDRLTEGLAEVGRGAAR  
SPARGFNREKVPVGIWEWPHRSGDSITLKAQVSPANPGFGLPTIGTVARFDD  
TTGSLTALIDGVLTAIRTGASAFASRLAAPDSVVLGSGQSVTQLHALTRV  
EVRVLTWDADPSHAASLADRVAFGLQVQAEPRHIAEADIICTATVPVGGPVL  
PADAVRPHLHINA:GRDLVGRKTEFLDLRLRAHVSPPHPEQARHEGECQQLPEPHIG  
LSLGLCADPREPTLCAAD"  
4994..6202  
/gene="visD"





KVLPSPAVIGLILVYCKVDVSLSGSHGIPELVCAAAVVLHLWKOMLLSIAGGTGVV  
KVLVQLVF

ORIGIN		Query Match	13.6%	Score 162.4	DB 1	Length 5008
		Best Local Similarity	51.4%	Pred. No. 2.1e-35		
		Matches 463	Conservative 0	Mismatches 416	Indels 21	Gaps 3
QY	227	CTATTGGAAATTCATCAATTAACATGAGCCGCGGAGCATACAAAATCCGTTCACTCG 286				
DB	1463	CTTTAACAACAATATGCTTAATTCAGATCCCTGATCAACACCGCTTCGAACTCTTG 1522				
QY	287	TGAACAAAGCCTTACTCCCGCGTGTAGCAATGGAACCGAGCAATCAAGAAATCA 346				
DB	1523	TTCAAAAGCAATTTAGCGATCGGATGATTTGCGATTTGGAAGACAAATTCAGCATATCG 1582				
QY	347	CAGATGAATGATTTCAAAATTTCAAGGGGCGAGTGTGATTTGACCTTSTTCACGATTTT 406				
DB	1583	CAGATCTTTTATTTGATTAAGTGCAGCCCAACAAATTCATGACCTAGTAGAGGATATG 1642				
QY	407	CATACCCGCTTCCGGTTTATTTGATATCTGAGTGTGCGAGTGCCTTCAGCGGAGATGG 466				
DB	1643	CGTTTCCATTCGCGATTTATCGTGAATGTAATGCTCGGCAATTCCTTCGGAAGATCGTC 1702				
QY	467	AACAGTTTAAAGCATGCTGTGATCTTCTGTCAGTACACCGAGGATAAAAGTCAAGAA 526				
DB	1703	AAAAATTCAGATGATGCTGCGA-----AGCAATTTGATTTTTCAGATGCCCC 1750				
QY	527	CTGAAAAGCCTTTTGGAAAGACGAGATAAGTGTGAGGAAGAACTGCGCGCTTTTGTG 586				
DB	1751	CTGAAAGA-----CTTCAGGAGATGACCACTTTTGTAGGGAGCTTTTGTGATATCTGG 1804				
QY	587	CCGGCATCATAGAAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTTTAGTGG 646				
DB	1805	AATCCTTAGTTTCGCAAAAGAGAGAGAACTCGCAGGAGATTTAATCAGTCGATTAATTC 1864				
QY	647	AAGCGGAAGAACAGGCGGAGAGAGTGTCCGCTGAGAGAGTCACTCCGTTTTCGACGCTGC 706				
DB	1865	AAGCGGAAGTGAAGGCGACACATTAAGTACAGAGAAATGTTACTCGATGATTATGCTGC 1924				
QY	707	TGCTGTGGCGGAAATGAAACCACTACAAACCTGATTTCAAATCGGATGACAGCATAT 766				
DB	1925	TTATCGTTGAGGTCATGAACACACCGTAACTTAATCAACCAATATGACATATGCTCTGA 1984				
QY	767	TAGAAACCCAGGCGTTTACGAGGAACTGCGGAGCCATCTGAACTGATCTCAGCAG 826				
DB	1985	TGTGTCAACATGATCACTAGAAAAGTCTCCGCGAGCTTGATGAATTCAGGCA 2044				
QY	827	TGGAGGAAGCCTTGGCTTTTTCAGAGCGCGCGCGCTTTTTCAGGGCGCATTCGCCAGCGGG 886				
DB	2045	TAGAGGAAGCATTCGGTTTTCAGAGCCTTGTGAGCTAACGACAAATTCGTCGAGCGCTG 2104				
QY	887	ATACGAGATCGGG---GGGCACTGTATTAAGAAGGTGATGTTTGGCGTTTGTGG 943				
DB	2105	AACCGTTTATCTGATCGCGCAGGAGATTAACGGAAGATGTGATCATCATTTCTTTAG 2164				
QY	944	CATCGGAAATTCGTGATGAAGCAAGTTTGAACAGCCGACATGTTTCATATCCGCCGCC 1003				
DB	2165	CATCAGCAACCGCTGATGAAAAGATTTTCCAAATTCGCGACATATTTGATATTTGAAGAA 2224				
QY	1004	ATCCCAATTCGCGATATTCGGTTTTCGCGAGCATCCATTTTTCCTTCGCGCGCGCTTG 1063				
DB	2225	AGAACACCGGCATATTCCTTTTGGCCCATGGCAATCATTTTGTCTTGTGTCAGAGTAG 2284				
QY	1064	CCCGTCTTGAAGCAATATCGGTTTAACTCTTTGATTTCTGCTTTTCTCATATGAGT 1123				
DB	2285	CCCGCTTAGAGCAAGATTCGATTTCCACCTTATAGCGCGTGTCCAAATATACAGT 2344				

RESULT 11  
BSU93876  
LOCUS  
DEFINITION Bacillus subtilis aminoglycoside 6-adenylyltransferase (aadK) gene,  
BSU93876 17828 bp DNA linear BCT 12-OCT-1997

```

/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="CAA71937.1"
/db_xref="GI:1926278"
/db_xref="GOA:008469"
/db_xref="SWISS-PROT:008469"
/translation="MSSEKKSIVTILTESQLSRAFKDRAEFYKELRSQALYPLSL
GALGKGLSRDADHLKNEKLLKKNYENYFAKEKRALLNKNEELTKHMLNSDPP
DNRRLTLVOKAFTHRMILQLEDKIOHIAEDSLDKQPKFNMVDDYAFPLIIVIS
EMLGIPLDEQRFVWSQAIIDFSDAPERLQENDHLLGFBFVLESLVRKQKREPAGD
LTSALQASSEGQSLSTELYSIMILLIYAGHETTVNLITNTYALMCHDDQLEKURQ
QPLMNSATEBALRPHSPVELTTRWTAPFPIIHGQEIIRKQDVIILSLASANDKIF
PNADIPDIRKKNRHAFGHGHEFCLAGAQLARLEAXIALSTLLRRCPNIQLKGEKQM
KWNGLFMRALDELPISF"
2696..2971
/notes="ORF91"
/codon_start=1
/transl_table=11
/product="hypothetical barstar-like protein"
/protein_id="CAA71938.1"
/db_xref="GI:1926279"
/db_xref="GOA:007938"
/db_xref="SPTREMBL:007938"
/translation="MRKIIIDGRDFENIEVLHDDKDKLIDPDYVYGRNLDALWDCLTG
WDLPLTLVLKNEFSNFTLGSVADDVLEVIQEAQELKDFKIIIE"
3305..3778
/gene="azlB"
3305..3778
/gene="azlB"
/functions="transcriptional regulator of the Lrp/AsnC
family; repressor of the azlBCDEF operon"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="transcriptional regulator"
/protein_id="CAA71939.1"
/db_xref="GI:1926280"
/db_xref="GOA:007920"
/db_xref="SWISS-PROT:007920"
/translation="MGIMLDETOKAILRDIQEDASINLNLKSKIGLSPSACLARTQN
LVGAGIITKETTIVDEKLGLEVTALALINLPLNRTHISFLEDINKFPQVQBCYTL
TGSHDYMLKIVAKDMESYRNFIIDSLMNPASIGVDTISVMSTKRTVSPIDEM"
3791..4555
/gene="azlC"
3791..4555
/gene="azlC"
/notes="product similar to Haemophilus influenzae
hypothetical protein; HI1738"
/codon_start=1
/transl_table=11
/protein_id="CAA71940.1"
/db_xref="GI:1926281"
/db_xref="GOA:007942"
/db_xref="SWISS-PROT:007942"
/translation="MWNKESLSLSSPAINTHMKRSQIWAAFRSAPFYTIPIFAGEL
FLGTAYGIPMSLGSFATYPIIMSGMIFAGSMFVAANFLGAFNPNALFLTLMVNA
RHLFGISMLDKYRGTKKLLYIFGCDSEFSINYTNVNPANVDKGMFEVTLNHH
LYWVAGAAIGGIFGYSVKFNTEGLDFVNTALFIVIFIQWMEKKHYSALTGLGLSVA
SLILFGNQOPIIPAMLAIGLVTLVRKPLEKAESV"
4552..4884
/gene="azlD"
4552..4884
/gene="azlD"
/notes="product similar to Haemophilus influenzae
hypothetical protein; HI1737"
/codon_start=1
/transl_table=11
/protein_id="CAA71941.1"
/db_xref="GI:1926282"
/db_xref="GOA:007923"
/db_xref="SWISS-PROT:007923"
/translation="MTMTMQMITIAMVLGTMLTRFLPFMFPSGKPTPKYQVLG

```

```

partial cds, and YrdA (yrdA), YrdB (yrdB), hypothetical protein
YrdC (yrdC), YrdD (yrdD), hypothetical cytochrome P450 protein YrdE
(yrdE), ribonuclease inhibitor YrdF (yrdF), regulatory protein YrdG
(yrdG), hypothetical protein YrdH (yrdH), hypothetical protein YrdI
(yrdI), amino acid transporter YrdJ (yrdJ), YrdK (yrdK), LysR family
regulatory protein YrdL (yrdL), YrdN (yrdN), cation transport
protein YrdO (yrdO), hypothetical protein YrdP (yrdP), lysR family
transcription regulator YrdQ (yrdQ), hypothetical protein YrdR
(yrdR) and hypothetical protein YrKA (yrkA) genes, complete cds.
ACCESSION U33876
VERSION U33876.1 GI:1934641
KEYWORDS
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
REFERENCE 1 (bases 1 to 17828)
AUTHORS Sorokin,A., Bolotin,A., Purnelle,B., Hilbert,H., Lauber,J.,
Dusterhoft,A. and Ehrlich,S.D.
TITLE Sequence of the Bacillus subtilis genome region in the vicinity of
the lev operon reveals two new extracytoplasmic function RNA
polymerase sigma factors SigV and SigZ
JOURNAL Microbiology 143 (Pt 9), 2939-2943 (1997)
MEDLINE 97453473
PUBMED 9308178
REFERENCE 2 (bases 1 to 17828)
AUTHORS Hilbert,H., Lauber,J., Bolotin,A., Sorokin,A., Dusterhoft,A. and
Ehrlich,S.D.
TITLE Direct Submission
JOURNAL Submitted (12-04-1997) Genetique Microbienne, CRJ INRA, Domaine de
Vilvert, Jouy en Josas 78352, France
FEATURES
source
1. .17828
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="1168"
/db_xref="taxon:1423"
/map="235 degrees"
1. .787
/gene="aadK"
/gene="aadK"
/gene="aadK"
/notes="similar to S.aureus plasmid pS194 Str protein"
/codon_start=2
/transl_table=11
/product="aminoglycoside 6-adenylyltransferase"
/protein_id="AAB80893.1"
/db_xref="GI:1934642"
/translation="VTLEGRSTNRNIPDPNDQYDISYFTDVSFKENDQWLEIFCK
RIMQGPEDMELPPELGNWFSYIILEPDGKLDLTLPIREAEYFANNQGLVKVL
DKDSFINYKTPNDQYWKRPAREFDDCCNEFMVSTTVVKGARNEILFALDHLN
EIVRNLRRMAWHASQKGYSTFSGMKYKPKRYLSNKWEELMSTYSVNGYQSMWK
SLFTCYALFKYSKAVSEGLAYKYPDYDEGITRYTEGIYGSVK"
1013. .1516
/gene="yrdA"
/gene="yrdA"
1013. .1516
/gene="yrdA"
/codon_start=1
/transl_table=11
/product="yrdA"
/protein_id="AAB80894.1"
/db_xref="GI:1934643"
/translation="MFQTLDFHLKSWFEADATQKLNLSITDRSLKQEIITSQNWTLGR
IAWHTVAAIGIITSNTDLTQAPAEYDVPVTSQFIADSYHQASNAFVQALKTWTH
TQGRINFIQGPNGSLMLFELIQHNHRRGQTVLMRQAGLTVPGIYGAKEWAKP
GLEAPKM"
2326. .2697
/gene="yrdB"
/gene="yrdB"
2326. .2697
/gene="yrdB"
/codon_start=1
/transl_table=11
/product="yrdB"
/protein_id="AAB80895.1"
/db_xref="GI:1934644"
/translation="MEKLNQTNLLIRFTLEIAALISGVYAMISPNCEYKXVLTLLVP
IACIIVMSFAVDFPSRSGQTVIANGVTRLVIELLIFAMAVALLFSIKPVSIVF
LCIILHIIISAERIKNLLNQ"
2709. .2728
/notes="transcription terminator"
2926. .3489
/gene="yrdC"
2926. .3489
/gene="yrdC"
/notes="similar to Arthrobacter sp. N-carbamoylsercoseine
amidohydrolase"
/codon_start=1
/transl_table=11
/product="hypothetical protein YrdC"
/protein_id="AAB80896.1"
/db_xref="GI:1934645"
/translation="MEBKALIIVDVQKAFDDKQKGGNNVKAENISKILEWRKRG
WTVIYQHTSDKPHSLPHKQEGFAIKIYKPMDEEVIITKNVSSFTGNLEFLKL
NEITVITGLTTPHCVSTTTRMSGNLGFDYTLISDATAAFGMEDQNDTYDAATIH
ISATLHDEFATLITDQLINDFKIH"
3588. .3737
/gene="yrdD"
3588. .3737
/gene="yrdD"
/codon_start=1
/transl_table=11
/product="yrdD"
/protein_id="AAB80897.1"
/db_xref="GI:1934646"
/translation="MFYQCARADEVSKVAKLHQSFNQYPFMDLLVREKKIHLFL
KCIKY"
4130. .5362
/gene="yrdE"
4130. .5362
/gene="yrdE"
/notes="similar to Saccharopolyspora erythraea cytochrome
P450CvIIB1"
/codon_start=1
/transl_table=11
/product="hypothetical cytochrome P450 protein YrdE"
/protein_id="AAB80898.1"
/db_xref="GI:1934647"
/translation="MSSKEKSVTILTESQLSSPAFXDEAFYFVKELRKSQALYELSL
GALGKWLISRYDDAIHLKNEKKNYENFJAKKRPALLKNEELTKMLNSDPP
DHNRLTIVQKAFTHRMILQEDKIQHIADSLLEKQPNKFMNLYDDTAFPLPVIS
EMLGIPIEDQKQFVMSQAIIDFSAERLQENDHLLGSFVEYLESVLRKKRREPAG
LISALIQASEGTOLSTELYMLMLLVAGHETTVNLTNTMTVLCMCHDDLEKLKQ
QPOLMNSAIEALRFHSPVELTTRMTAEPLFHQEIKKKDVIIISLASANRDEKIF
PNADIPDIERKNNRHIAFGHGHFCLGQALRLAKIAISTLLRRCNPVQLKGEKQM
KWKGNFLMRALDELPISF"
5634. .5909
/gene="yrdF"
5634. .5909
/gene="yrdF"
/notes="similar to Bacillus amyloliquefacience ribonuclease
inhibitor (barnstar)"
/codon_start=1
/transl_table=11
/product="ribonuclease inhibitor"
/protein_id="AAB80899.1"
/db_xref="GI:1934648"
/translation="WRKIIIDGRDFENIEVHDDLKDLKDLPPDYGRNLDALDCLTG
WVDPLPLTLKNFEFSNTFLGSYADTVLEVIQEAQELKDEFFKIIIE"
6243. .6716
/gene="yrdG"
6243. .6716
/gene="yrdG"
/notes="similar to E.coli leucine-responsive regulatory
protein"
/codon_start=1
/transl_table=11
/product="regulatory protein YrdG"

```

```

/protein id="AAB80900.1"
/db_xref="GI:1934649"
/translation="MGIMLDETDKAILRLDQEDASISNLNLSKIGLSPSACIARTKY
LYEAGLIKPTTIVDEKLGIVTALALINLSPNRETHSFLIEDINKPPOVECVTL
TQSHYMLKIVADNMEYSYRNFIIDSLMOPALISGVDTSIVMTEKTSVPIDEM"
6729..7493
/gene="yrdH"
CDS
6729..7493
/gene="yrdH"
/notes="similar to H.influenzae hypothetical protein
H11738"
/codon_start=1
/transl_table=11
/product="hypothetical protein YrdH"
/protein id="AAB80901.1"
/db_xref="GI:1934650"
/translation="MKNKESLHSSPAINTHMNKSQIWAAPSPAPVPIPIFAGFL
FLGIAYGIEMHSGFSATPIPIIMFSPILFAGSMHFVAANFLGAPNEMNALFLTLVNA
RLHYGISMLDKYRGTKKKLYLIFGMCDESPSINTANVPANVKGWFMFFVILLNH
LYWVAGAAIGGIFSVYKFNTEGLDFVMTALFVIEIQWMKEKKHYSALTIGLSVA
SLILFGNFIIPAMLAALLVLRKPLEAEVSV"
7490..7822
/gene="yrdI"
CDS
7490..7822
/gene="yrdI"
/notes="similar to H.influenzae hypothetical protein
H11737"
/codon_start=1
/transl_table=11
/product="hypothetical protein YrdI"
/protein id="AAB80902.1"
/db_xref="GI:1934651"
/translation="MTMTQOMITIAMVVLGMLTRFLPFMIFPSKPKPKYVQVYG
KVLPSAVIGLLVIYCLKDVSLLSGSHGIPELVGAADVLLHLNKKNNLLSIAGTVVY
MWLVQLVF"
7833..7856
/notes="transcription terminator"
7987..9309
/gene="yrdJ"
CDS
7987..9309
/gene="yrdJ"
/notes="similar to branched chain amino acid carrier BrnQ
from Lactobacillus delbrueckii"
/codon_start=1
/transl_table=11
/product="amino acid transporter"
/protein id="AAB80903.1"
/db_xref="GI:1934652"
/translation="MSKKVSASYIIILGLMLFALFCAGNLIPPPMLGQLAGKNVVA
NAGFLVTGVLPLAITAFVSGKQNSLASRHPVFGIVFTILILALGPFAIPR
SGNVSEIGVKPLISNDASPSGLIIFTLFLALCLSLNPSKIIDIVGKFLPIKLT
FIGLVAVALIRPIGTITQAPSKGVTQAPFKQEGYLTDLALVAFVFGHIIIVNALKE

```

```

Query Match 13.68; Score 162.4; DB 1; Length 17828;
Best Local Similarity 51.48; Pred No. 2.7e-35;
Matches 463; Conservative 0; Mismatches 416; Indels 21; Gaps 3;

QY 227 CTATTGAAATTCATCATTAACATGGACCGCGAAGCATACAAAATCGTTGAGTCG 286
DB 4401 CTTTACAAACATATCGTTAATTCAGATCCCTGATCACACCGCTTGGAATCTTG 4460

QY 287 TGAACAAAGCGTTTACTCGCGCGTGTGAAGAGCATGGGAACCGAGATTCAGAAATCA 346
DB 4461 TTCAAAAAGCATTTACGCATCGGATGATTTTSCAGTTGGAAGACAAAATCCAGCATATCG 4520

QY 347 CAGATGAACTGATCAAAAATTCAGGGCGCAGTGTGAGCTTGCTTTCACGATTTT 406
DB 4521 CAGATTTCTTATTCGATAAAGTCAGGCCACCAAAATTCATGAACCTAGTACGAGATTAAG 4580

QY 407 CATACCCGCTTCCGGTATTCTGTATCTGAGCTGCTGGGAGTGCCTTTCAGCGAGATGG 466
DB 4581 CGTTTCCATTGCCGATATCTGTATAGTGAATGCTCGGATTCCTTCTTGAAGATGCTC 4640

```

```

QY 467 ACAGTTTAAAGCATCGTCGATCTTCTGTGTAGTACACCGAAGGATAAAAGTGAAGAG 526
DB 4641 AAAAATTCAGAGTATGGTCGCA-----AGCAATTTGATTTTTCAGATGCC 4688
QY 527 CTGAAAAAGCGCTTTTGGGAAGAACGAGATAAGTGTGAGGAAGAACTGGCCGCTTTTGG 586
DB 4689 CTGAAAGA-----CTTCAGGAGATGATCCACCTGTGTAGGGAGTTTGTGTAATATCTGG 4742
QY 587 CCGGCATCATAGAAAGAAACGAGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGG 646
DB 4743 ATCTCTTAGTTTCGAAAAAAGGAGAGAACCTCAGGAGATTTAATCAGTGCATTAATTC 4802
QY 647 AAGCGGAAGAAACAGCGGAGAGAGTGTCCGTTGAGAGCTGATTCGCTTTTTCGACGCTGC 706
DB 4803 AAGCGGAAGTGAAGGACACAAATTAAGTACAGAGAAATGTACTGATGATATTATGCTGC 4862
QY 707 TGTGTGGTGGCCGAAATGAAACCACTACAAACCTGATTTTCAAAATGCGATGTACAGCAAT 766
DB 4863 TTATCGTTGAGGTCTATGAAACCAACCGTAACTTAATCAACAATATGACATATGCTCTGA 4922
QY 767 TAGAAACCGCAGCGCTTTAGCAGGAACTGGCAGCCATCTGAACTGATGATGCTCAGCAG 826
DB 4923 TGTGTCAACATGATCAACTAGAAAAGCTCGCCAGCAGCCGAGCTTGTGAATTCAGCCA 4982
QY 827 TGGAGGAAGCTTTCGCTTTTCAGAGCGCGCCCGCTTTTGGAGCGCATTCGCAAGCGGG 886
DB 4983 TAGAAGAGCATTTGCGGTTTTCAGAGCCCTGTTGAGCTTAAACGACATTCGCTGGAGCGCTG 5042
QY 887 ATACGAGAGATCGGG---GGGCACCTGATTAAGAGGTGATATGTTTGGCGTTTGTGG 943
DB 5043 AACCGTTTACTGATCGATGGCAGGAGATTAACGGAAGATGTGATCATATTTCTTTAG 5102
QY 944 CATCGGCAAAATCGTGTATGAAAGCAAGTTTACAGACCGCAGCATGTTTGTATTCGCGGCC 1003
DB 5103 CATCAGCAAAACCGTGTATGAAAGATATTTTCCAAATCGGCACATATTTGATTAAGAAAG 5162
QY 1004 ATCCCAATCCGATATTCGCTTTGCGCAGCGCATCTTCATTTTTCCTTGGGCGCCGCTTG 1063
DB 5163 AGAACAAACCGCATTTGCTTTTGGCCATGGCAATCATTTTGTCTGTCACAGCTAG 5222
QY 1064 CCCGCTTGAAGCAAAATATCGCTTAAAGCTTTTGAATTTCTGCTTTTCTCTCATATGAGT 1123
DB 5223 CCCGCTTAGAGCAAAAGATGCGATTTCCACCTTTATAGCGCGCTGTCCAAATATACAGT 5282

```

```

RESULT 12
BSUB0014/c 199173 bp DNA linear BCT 07-JUL-2003
LOCUS Bacillus subtilis complete genome (section 14 of 21): from 2613658
to 2812830.
DEFINITION Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
ACCESSION Z99117 ALJ009126
VERSION Z99117.2 GI:32468796
SOURCE
ORGANISM
REFERENCE
AUTHORS

```

```

Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 199173)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borcher, S.,
Borriss, R., Bourcier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entlan, K.D., Errington, J., Fabret, C., Ferrati, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,
Henau, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koeter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,

```

Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Prescann, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, P., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

# **TITLE**

The complete genome sequence of the gram-positive bacterium

*Bacillus subtilis*

Nature 390 (6657), 249-256 (1997)

93844033

9384377

2 (bases 1 to 199173)

Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

Direct Submission

Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,

Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724

Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,

adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45

68 89 48

On Jul 7, 2003 this sequence version replaced gi:2634966.

This entry contains data from release R16.1 of the Subtilist

database. Further data on gene annotation and detailed information

about changes from previous releases can be found at

<http://genolist.pasteur.fr/Subtilist/>.

Location/Qualifiers

1..199173

organism="Bacillus subtilis subsp. subtilis str. 168"

/mol\_type="genomic DNA"

/strain="168"

/db\_xref="taxon:224308"

complement(71..1030)

/gene="phoH"

/locus\_tag="BSU25340"

complement(71..1030)

/gene="phoH"

/locus\_tag="BSU25340"

/note="alternate gene name: yqfB"

/codon\_start=1

/transl\_table=11

/product="phosphate starvation-induced protein"

/protein\_id="CAB14476.1"

/db\_xref="GI:2634980"

/db\_xref="GOA:P46343"

/db\_xref="SWISS-PROT:P46343"

/translation="MTEHLAMQKKNPDEALSLFGNODSLKLMKDLNLITRG

ETIVSGDSQFADLLGSLALIRKGIETSERDVIYAIKWAKNLEVPESMVEE

EITNAKSKIRVTKQREYVAMKNDLPVIGIPAGCTGYLAVKVAHLKNGHI

KXILITPAVAGESLPGDLKEKVDYPLRPLDHLVGDAGHDSRLMERGLIEI

APLPMRGRTLDAYILDHQAQNTTPAQMKPFLTFSPSSKMIITGDSIDLPKGVK

SLAVAKMLKIGISIMIELDQTDVVRHPLVAKIIIEYDKQ"

complement(1027..2223)

/gene="yqfD"

/locus\_tag="BSU25350"

complement(1027..2223)

/gene="yqfD"

/locus\_tag="BSU25350"

/function="unknown"

/note="similar to stage IV sporulation protein"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB14477.1"

/db\_xref="GI:2634981"

/db\_xref="GOA:P54469"

/db\_xref="SWISS-PROT:P54469"

/translation="MKRWLSFFSGKVQLTGRGIERLLNCTKGIPVHRKKKKE

AVSLIQLQDVHAFRRVSKPKKARFINRKGFPFLLLKSKLNIGFTIGFAIFPILFLF  
LUSNMVKLDITGAKPDETEHQHROHNLNIGVKKGRLOPLMSPKIQKSLTWGIDKNT  
WYDILKGTTHMKVVERNEPEKIVSPRNIVAKKAILITRMSVQGGQFMAIHVHV  
KGGQLLVGLIGSDHQDEKASKAIYGETWYRSEVTVPLETLENVYTSKRTKKLS  
FGSLAIPWGMTFKKEELKHPTEQEKSHLFLGFKLPVSYVKEQTRSEEARLKYTK  
BEAVQGIKLGKQDVDEKIGENGEVSEKVLHQTVENGKVLIIYQVIEDIVQTPI  
VRETE"

complement(2242..2523);  
/gene="yqfC"  
/locus\_tag="BSU25360"  
complement(2242..2523);  
/gene="yqfC"  
/locus\_tag="BSU25360"  
/function="unknown"  
/note="similar to unknown proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14478.1"  
/db\_xref="GI:2634982"  
/db\_xref="SWISS-PROT:P54468"  
/translation="MQQRNRMKAWLTRALEIPDPVMDLPRTVMGRLHIYENRG  
LILFSEVRLMKQGCIIISGNFVIKAILPEELLEGITDVRYVES"  
complement(2580..2999)  
/gene="yqfB"  
/locus\_tag="BSU25370"  
complement(2580..2999)  
/gene="yqfB"  
/locus\_tag="BSU25370"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14479.1"  
/db\_xref="GI:2634983"  
/db\_xref="GOA:P54467"  
/db\_xref="SWISS-PROT:P54467"  
/translation="MEDLLTNLIILAIIGISAIIGKSKKEKONSQKRRKPKHVQS  
ASPKQKSKEDAPAPINRMVQAREASERRETRNLKGLERDLAAAKQKTVYTKQK  
MLQVNRKLVVQGVILGEVFGPPRAKKPHRTMPARK"  
complement(3024..4019)  
/gene="yqfA"  
/locus\_tag="BSU25380"  
complement(3024..4019)  
/gene="yqfA"  
/locus\_tag="BSU25380"  
/function="unknown"  
/note="similar to unknown proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14480.1"  
/db\_xref="GI:2634984"  
/db\_xref="GOA:P54466"  
/db\_xref="SWISS-PROT:P54466"  
/translation="MDPSTMIILAVAVAILVAVFTFVFMVLMISALAAGVKISIF  
TLVGRRLRVIPNRVNPILIKAKAGLVNQLSHYLAGNVORVVALIAAQRAN  
IELTERCAAIDLAGRDVLEAVQMSVNPKEVETPFITAGVMDGIEVKAKRITVRA  
ERLVGAGEETIVAVUGGIVSTIGSSDNHKKVLENPDMTSQTLGKGLDSTAPAIL  
SIDTADVGITGNIGALLQDQADKGNIAQAKERRAMAVAOEQEMRAREMRAKV  
VBAEAEVPLAAEALREGNIGVYMNINIKNIDATEMRDPSFGKLTDPDSEDRKS"  
complement(4041..5354)  
/gene="yqez"  
/locus\_tag="BSU25390"  
complement(4041..5354)  
/gene="yqez"  
/locus\_tag="BSU25390"  
/function="unknown"  
/note="similar to unknown proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14481.1"  
/db\_xref="GI:2634985"  
/db\_xref="GOA:P54465"  
/db\_xref="SWISS-PROT:P54465"  
/translation="MLQIKGFRAALLGIFLLSILGVQLNAKAEKQTVYVIVPEKNVEQ

GLASFLSRSLQDAKADAHADHIIIDINTPGGLVKSADIMADLI TESELPVTA VYVKRAL  
 SAGAYTALQADHYIYAPGGKMGAAAI VDGQNAADQKAOQLWLAEMDEDAVKNRDPK  
 YALAMADPDIDAKEVAPKGLDILNADKAI EVGSEGTADNLSTLVKLI GFPERAQIS  
 YAKESFAEKTRWLTNVIPIVILITIAFLGLITVELSPGVCLPGTAGLIALILFFYGH  
 LAAGLAGYEVLLPIAGVILILILEIPUPGGIIGLIGLAIIASLFLAAGSFTVMVAVSL  
 LIASAVSITAFILITLTVLGRMKPFKKLILNDSTNTESGVSVNQTDRDLMKVGLITF  
 PLPSGTVIDDRLDVSRSSTFEKDKKKVVKVIVKESRIVVREI"  
 complement (5485..5931)  
 /gene="ygey"  
 /locus\_tag="BSU25400"  
 complement (5485..5931)  
 /gene="ygey"  
 /locus\_tag="BSU25400"  
 /functions="unknown"  
 /note="similar to unknown proteins"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB14482.1"  
 /db\_xref="GI:2634986"  
 /db\_xref="SWISS-PROT:P54464"  
 /translation="MSLLRLNQMKLYMKNREKDKLTVVVMVKASLQNRKILKNDK  
 LTEDLELTLSLRKQKDSLOEFSNARLDLDVKQKELDILEVYLPEQLSREELT  
 IVNETAEVGSASKADMGKVMGAI MPKVKGADGSLINKLVSSQLS"  
 complement (5946..6119)  
 /gene="rpsU"  
 /locus\_tag="BSU25410"  
 complement (5946..6119)  
 /gene="rpsU"

Query Match 13.68; Score 162.4; DB 1; Length 199173;  
 Best Local Similarity 51.48; Pred. No. 4.6e-35;  
 Matches 463; Conservative 0; Mismatches 416; Indels 21; Gaps 3;

QY 227 CTATTGGAATTCATCATTAACATGACCGCGCGAGCATACAAAATCGTTCACTG 286  
 DB 117637 CTTTAAACAAACATATGCTTAATTCAGATCCCCCTGATCAACCGCTTGGAACTCTTG 117578

QY 287 TGAACAAAGCTTTACTCCCGCGGTGATGAGCATGGGACGAGATTCAGAAATCA 346  
 DB 117577 TCCAAAAGCATTTTACGCATCGGATGATTTTGCAGTTTGGAGACAAAATCAGCATATCG 117518

QY 347 CAGATGAACCTGATTCAAAAATTTCCAGCGCGCGAGTGGATTGACCTTGTTCCACATTTT 406  
 DB 117517 CAGATCTTTATTGGATTAAGTGCAGCGCCACCAAAATTCATGACCTAGTAGACATTAIG 117458

QY 407 CATACCGCTTCCGGTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGG 466  
 DB 117457 CGTTTCCATTCGCGGATTCGTGATAAGTGAATGCTCGGCAATTCCTCTTGAAGATCGTC 117398

QY 467 AACAGTTTAAGCATGCTGTGATCTTCGTGTAGTACACCGAGGATAAAGTGAAGAG 526  
 DB 117397 AAAAAATTCAGAGTATGTCGCA-----AGCAATTAATGATTTTTCAGATGCC 117350

QY 527 CTGAAAAAGCCTTTTGGAAAGACGAGATAAGTGTGAGGAAGACCTGCGCGCTTTTGG 586  
 DB 117349 CTGAAAGA-----CTTCAGAGGAATGACCACTTGTAGGGAGTTTGTGAATATCTGG 117296

QY 587 CCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATTATTTCTATTATTAGTG 646  
 DB 117295 AATCTCTAGTTTGGCAAAAAGGAGAGAACCTGCGAGGAGATTTAATCAGTGCATTAATTC 117236

QY 647 AGCGGAGAGAAACAGCGAGAGTGTCCGCTGAGAGCTGATTCGCTTTTGGACGTGC 706  
 DB 117235 AGCGGGAAGTGAAGGACACCAATTAAGTACAGAGAATTTGACTCGATGATTAATGCTGC 117176

QY 707 TGTGTTGGCGCGGAAATGAACACCACTACAAACCTGATTTCAAATGCGATGACAGCATAT 766  
 DB 117175 TTATCGTTGAGGTCATGAACACACCGTAACTTAATCAACCAATATGACATATGCTCTGA 117116

QY 767 TAGAAACCGCAGCGGTTTACGAGGAATGCGGACGCCATCTCGAACTGACCTCAGCGAG 826  
 DB 117115 TGTGTACCATGATCAACTAGAAAAAGCTCCGCCAGCAGCGGCTTGATGAATTCAGGCA 117056

QY 827 TGGAGGAAGCCTTTCGCTTTAGAGCGCGCGCGCTTTTGGAGCGCATTCGCAAGCGG 886  
 DB 117055 TAGAAGAAGCATTCGCGTTTACAGCCCTGTTGAGCTTAACCAATTCGCTGGACGGCTG 116996

QY 887 ATACGGAGATCGGG---GGGCACCTGATTAAGAAGAGTGATGTTTGGCGTTTGTGG 943  
 DB 116995 AACGTTTATACTGCATGGCGGAGAGATTAACCGGAAGAATGTCATCATCATTTCTTAG 116936

QY 944 CATCGCAAAATCGTGATGAAGCAAAAGTTTGACAGACCGCATGTTTGATATCGCCGCC 1003  
 DB 116935 CATCAGCAAAACCGTGATGCAAAAAGATATTTCAAAATGCGCATATTTGATATGAAGAA 116876

QY 1004 ATCCCAATCCGATATTGCGTTTGGCAACGCATCCATTTTGGCTTGGGCGCGCTTG 1063  
 DB 116875 AGAAACAACCGCATATTGCTTTTGGCCATGCAATCATTTTGTCTTGGTGACAGCTAG 116816

QY 1064 CCGCTCTTGAAGCAATATCGCTTAACGCTTTTGAATTTCTGCTTTTCTCATATGAGT 1123  
 DB 116815 CCGCTTAGAAGCAAAAGATTCGGAITTCACCTTATACGCCGCTGTCCCAATATACAGT 116756

RESULT 13  
 AE017006  
 LOCUS 300893 bp DNA linear BCT 16-MAY-2003  
 DEFINITION Bacillus cereus ATCC 14579 section 9 of 18 of the complete genome.  
 ACCESSION AE017006 AE016877  
 VERSION AE017006.1 GI:29896139  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bacillus cereus ATCC 14579  
 Bacillus cereus ATCC 14579  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus  
 cereus group.  
 REFERENCE  
 1 (bases 1 to 300893)  
 Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelson,B.,  
 Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N.,  
 Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,  
 Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M.,  
 Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.  
 Genome sequence of Bacillus cereus and comparative analysis with  
 Bacillus anthracis  
 Nature 423 (6935), 87-91 (2003)  
 JOURNAL  
 MEDLINE 22608415  
 PUBMED 12721630  
 REFERENCE  
 2 (bases 1 to 300893)  
 Candelson,B., Galloux,K., Ehrlich,D.S. and Sorokin,A.  
 The number of ribosomal RNA operons in Bacillus cereus  
 Unpublished  
 JOURNAL  
 REFERENCE  
 3 (bases 1 to 300893)  
 Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelson,B.,  
 Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N.,  
 Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,  
 Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M.,  
 Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.  
 Direct Submission  
 Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de  
 Vilvert, Jouy en Josas 78352, France  
 JOURNAL  
 TITLE  
 FEATURES  
 source  
 join(1..105823,144296..300893)  
 /organism="Bacillus cereus ATCC 14579"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 14579"  
 /db\_xref="ATCC:14579"  
 /db\_xref="taxon:226900"  
 /focus  
 .05824..144295  
 /organism="Bacteriophage phiBC5A52"  
 /proviral  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:225726"  
 /note="identified by comparison with Bacillus anthracis"  
 192..896  
 /locus\_tag="BC2457"  
 192..896  
 CDS

/locus\_tag="BC2457"  
 /EC\_number="2.7.8.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="4'-phosphopantetheinyl transferase"  
 /protein\_id="AAP09420.1"  
 /db\_xref="GI:29896140"  
 /translation="MKIPKNTLSHLFKKLSNLVSGKKRMKRLNSCDINRTLIGDL  
 LIRSLICQYKINNEIRFLVNEYGKPVENFPHNLSSGEMVCTTANFNGID  
 IERVSEIRAFKLAKEFSAEAFYDLISNMSDEQINYFYDLWTLKESYIKTKIGKLYTP  
 LNSFSIKKESRLISYKHIPKPNFYKQYNIDPNTKLSACATRDFFPQBIINDIYTC  
 QTIYKFSKEKINAED"  
 883..1605  
 /locus\_tag="BC2458"  
 883..1605  
 /locus\_tag="BC2458"  
 /EC\_number="3.1.2.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Thioesterase"  
 /protein\_id="AAP09421.1"  
 /db\_xref="GI:29896141"  
 /translation="MQTKLFCFPHAGSASFYAKWKHFNPIVVPVELAGCYRI  
 SENYQSEERAVNDYNIKVMQIDDSYILFHSNGSLIAVEARKIKDSKNVSPEFL  
 VLSGRNHSKIRNIRHLSNEQFKREVIAMGGTSPGVQSEELMEI FLPIILRADPKI  
 VETYIHDKNTQPCDIDFLIPNGENDEFTYDQVWKERYTSKTCFHSFEGNHFFNE  
 NIEETANSIKRKLDSKRLNSF"  
 1798..1868  
 /locus\_tag="BC2459"  
 1798..1868  
 /locus\_tag="BC2459"  
 /product="tRNA-OTHER"  
 complement(3310..3900)  
 /locus\_tag="BC2460"  
 complement(3310..3900)  
 /locus\_tag="BC2460"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Succinodiglycan biosynthesis protein"  
 /protein\_id="AAP09422.1"  
 /db\_xref="GI:29896142"  
 /translation="MIATFETLLWEAADVQONFDNSTATAMQSLYSVWVTEELBQLF  
 TYMKQEKKEKPLTLAGEDMNLFYKSSFSRYAKWLQKYNPEVSEPTAVSELIELD  
 RYNNKNTYPPYDRFQWETQPVINKPKVRMFIQNHKSELQVTPHTYDVKVRGGISS  
 ENLTDYSPIYDQKQKQNFAMLTENQYGNKLAPFNK"  
 4397..5044  
 /locus\_tag="BC2461"  
 4397..5044  
 /locus\_tag="BC2461"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Antigen"  
 /protein\_id="AAP09423.1"  
 /db\_xref="GI:29896143"  
 /translation="MKDTSKKQIIKVFLLSLGLVLIILGLMYPNHNKNTIQNKAQATE  
 KRLVQYESTLKKELNIGKERTPILGIMYQESRGEQNDPMQSSSELGLKNEIQTET  
 SLISBOGVKHFARKYKGTQKDVSDTIQSYNMGSGYIDFVASQEVQKQSDSAKFF  
 SKMKYDQNPAMYTCGGNNKNFRYPYCYGDFTYATKVNKTKLIEELLRNHSSSK"  
 5362..5868  
 /locus\_tag="BC2462"  
 5362..5868  
 /locus\_tag="BC2462"  
 /EC\_number="2.3.1.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Ribosomal-protein-serine acetyltransferase"  
 /protein\_id="AAP09424.1"  
 /db\_xref="GI:29896144"  
 /translation="MFIKTERLLIRKFEKMQAVHEYTSIDINVMKYIPEGVTEBNT  
 RNFVNENIGENAKNFVVLHKDILLDHIVFHYKFGCHTYEIGWFPNPKYKNGYASE  
 AQAATLEYGFKEKHLRIITATQCPQNTSYRNVMEKIGMRREGYFRKCIPIHGBWWDYEY  
 YVAILEEE"  
 complement(6235..6420)

/locus\_tag="BC2463"  
 complement(6235..6420)  
 /locus\_tag="BC2463"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAP09425.1"  
 /db\_xref="GI:29896145"  
 /translation="MQYICIRKCIIVISAYTNKVIKIYIILIRKGENNMEIAVVLK  
 FIGNAIPIUQELLKVFEM"  
 7321..8772  
 /locus\_tag="BC2464"  
 7321..8772  
 /locus\_tag="BC2464"  
 /EC\_number="3.2.1.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="S-layer protein / Peptidoglycan  
 endo-beta-N-acetylglucosaminidase"  
 /protein\_id="AAP09426.1"  
 /db\_xref="GI:29896146"  
 /translation="MINTWVEVFLMKHKLJATGILAGAILSYSSNLAIAHTKFPDVP  
 TWADKSVNLVDKQVLNGYVDGTFGSNDLSDRASATKIMTKVLGKIDPNAKPSPTDS  
 QNHATPTIAAEKAGIVKSGNGIENPSGKYTRAAMATMLVNAKLGSTAHNEQVR  
 FEDLGHGKEGHANLIDLIKISNGHENGQPNRFTIRAEAGLTAKTMDLQONLDEK  
 EVTATSVEDNLITVASLITAEIDSTIAKHSDSPFVGHGGQFNAQNYGVSAHYL  
 RAHALLESYGKSEIAYQKHNLFGLRAYDGPFRYAKYLPSPGDSIYANVYRERYL  
 RESGMVYNGSTLTGMNVYASDKWAKIAGIMERIKPFRVEDYTYAKKLKPKNETLD  
 VDALSNEIPKMYADGSSSNVSSATYYQVPYFNLKIKSRPDVAVEENKVGTVTPGT  
 TIFIYREDPNGVFESFEANGKYYWTLKNKLSM"  
 9888..10085  
 /locus\_tag="BC2465"  
 9888..10085  
 /locus\_tag="BC2465"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAP09427.1"  
 /db\_xref="GI:29896147"  
 /translation="MISLGINILVPLSPFFIGGMATDPSGSTMDFEVEFLFIQIPFF  
 PLVLSLVMLVRRKKEKVVH"  
 11619..12941  
 /locus\_tag="BC2466"  
 11619..12941  
 /locus\_tag="BC2466"  
 /EC\_number="3.4.24.3"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Microbial collagenase"  
 /protein\_id="AAP09428.1"  
 /db\_xref="GI:29896148"  
 /translation="MYQNERLTWFQEGNAEFFAGSTRTNVVPKRSIIISGLSDPASR  
 YTAQOTFAKYGSDPYKSPALOSYLYNHQPTFDKQLCLIRANDVXNDVAYRELLS  
 KNTQNAEYQAYMOQLTDQGNVYHQVNDYLIQHPKPLAEKNEIYDVANIQDAK  
 ITKHSQFNTESEVGTYYGGSKESEDMKTSKOVNCTLSQSKQSGSVYTVAY  
 FVYRVLNMAFEYDILVFHGVATEEKEKTTIIVNMAGPSGSGIYNEIIOHNSGTTSEN  
 GKVIYLMNFGDGIITSKNPHITGYEKSTVELTVKDSRGKSEIQTKVTYKQDPQ  
 TGSILEEEKILPENTLAKNLTVDQTDVYTNVTDPKVEISVNVNEQNTGMTWLYH  
 ESDMNYVACGEDEGNTIKGKFAKPGKYLVNHFEDDKNGEYSLLVK"  
 13602..14639  
 /locus\_tag="BC2467"  
 13602..14639  
 /locus\_tag="BC2467"  
 /EC\_number="3.4.16.4"  
 /codon\_start=1  
 /transl\_table=11  
 /product="D-alanyl-D-alanine carboxypeptidase"  
 /protein\_id="AAP09429.1"  
 /db\_xref="GI:29896149"  
 /translation="MNNTYNEKTHTLIKOLLKFSKVPKPGFAYWASFDSGVTYKGAVG  
 LASIENLPIITKSVENIASVSKQFTAFSILLILOEGRSLSDDSIVKFPVSIQTYAEP  
 VTLRH:LIHHTGGLVDYMLAHAEINIEFTDILTVEQSLKHLNGHQIARFPVGTKEYSN  
 TGYFLLSLVVEKVSGLKQFAKERIFDSLNMKDDITIVDCYPTTIIISARGYSKNEQA

gene		Query Match		13.4%; Score 159.6; DB 1; Length 300893;	
		Best Local Similarity		51.6%; Pred. No. 3.2e-34;	
		Matches 453; Conservative		0; Mismatches 404; Indels 21; Gaps 3;	
YMYVSPWHTGDDGAVHATVEDLVKNGENLTGTGGKELVGRMEIGREISPTRETI					
IANEDYAFGLRLAEGNCRYLEHSGWAGSYRPFKEXYLTVVVLSNYDGFDSKKY					
ANEIAGILIEK"					
14889..16856					
QY	241	ATCATTAAATGATGAGCCGCGGAGGATACAAATAACCGTTAGTGTGATGACAAAGCCTTT	300		
DB	151930	ATGCTAAATTCAGATCCACCAACCAACGCTGTTAGCATCTTTAGTTCAAAAGCTTTT	151989		
QY	301	ACTCGCGCGTATGAAGCAATGGGAACCGAATTCAGAAATCAAGATCAAGATCAATGAT	360		
DB	151990	ACACCGAAGATGATTGCAATTTAGACGGAAGATTCAGAGAATAGCAGATGATTGATA	152049		
QY	361	CAAAAATTCAGGCGCGAGTGAAGTTGACCTTGTTCAGATTTTTCATACCGCTTCCG	420		
DB	152050	AGTGATATAGCGAAAGGATCAATTAATCTGTGGATGATTTCAITTCATTTCCATTACCA	152109		
QY	421	GTTATTGTGATATCTGAGCTGCTGGAGTGCCCTTCAGCGCAGATCGAACAGTTTAAAGCA	480		
DB	152110	ATTATGTAATAGCAGATGCTCGTATTCAAAGAGATCAAGCGAAATTTAGATTT	152169		
QY	481	TGGTCTGATCTTCTGTCAGTACCGAAGGATAAAAGTGAAGAGCTGAAAAGCCTTT	540		
DB	152170	TGGTCTCATGCT---GTTATTCATCACCAGAAACACCTGAAGAA-----	152211		
QY	541	TTGGAGACGAGATAAGTGTGAGGAGACTGCGCGCTTTTTCGCGGCATCATAGAA	600		
DB	152212	ATAAAGAGACTGAAACAACTATCTGAATTTATTAACATATCTCAATTTAGTTGAT	152271		
QY	601	GAAAGCGAAACAAACCGGACAGGATATTTCTATTTTAGTGAAGCGGAAGAAACA	660		
DB	152272	ATTAAGAAAGAGCCAAAGAGACTTGGTGAAGTCTTAACTTTCAGAGAGTGAA	152331		
QY	661	GCGGAGAGCTGTCGCGTGAAGAGCTGATTCGCTTTTGCAAGCTGCTGTTGGCGGAA	720		
DB	152332	GCGCAATAACTTAGCGCTCGGGAATATATTCAAATGATATCTTAATTTGTCGACGGA	152391		
QY	721	AATGAACCACTACAACTGATTTCAATGATGATGATGATGATGATGATGATGATGAT	780		
DB	152392	CATGAGACGACAGTGAATTTAATTAATCAAAATCGGTTATAGCACTTCTTCAAAATCCAAAT	152451		
QY	781	GTTTACGAGGACTGCGGACCATCTGAAGTATGCTTCAGCGAGTGGAGAGCCTTG	840		
DB	152452	CAATTCAGTTATTAAGATATCCAAACTATTTGATTCGGCTATTCAGGAGGATG	152511		
QY	841	CGTTTCAGAGCGCGCGCGCGCTTTGAGCGGATTCGCAAGCGGATAC---GGAGATC	897		
DB	152512	CGTTATTTATTCAGTTGAGGTTCAACTGCAAGATGGCGGACCTTTTCAATTT	152571		
QY	898	GGGGGCGACCTGATTAAGAGGTGATATGTTTGGCGCTTTGTCGATCGCAATCGT	957		
DB	152572	CACCATCAACAAATACAGAAAGAGATATGTTTATTTGATTCGATTCGCTTCAGGAAACCGT	152631		
QY	958	GATGAAGCAAGTTTGACAGACGACATGTTTGATATCGCGGCCATCCCAATCCGCAT	1017		
DB	152632	GATGAACAGATTTTGAATCCAGAAATATTTGATTTACAGCGGAGAACACCGTCAAC	152691		
QY	1018	ATTGCGTTTGGCAGCGCATTCATTTTTCGCTTTGGGGCGCGCTTCCCGCTTTGAAGCA	1077		
DB	152692	ATTGCGTTTGGTATGCTAGTCAATTTCTGCTAGGAGCTCCATTCGCGAGGTTAGAAGCA	152751		
QY	1078	ATATCGGTTTAACGCTTTTGTATTTCTGCTTTTCTCA 1115			
DB	152752	AAGATGCTTACTACTTCTTGTATTAATGCAATGCCCTGA 152789			
RESULT 14		SFU08223		10167 bp DNA linear BCT 04-FEB-2000	
LOCUS		SFU08223			

DEFINITION	Streptomyces fradiae cytochrome P-450, dTDP-glucose synthase, dTDP-glucose dehydratase, thioesterase, TylCVI (tylCVI), and TylR (tylR) genes, complete cds; and unknown gene.
ACCESSION	U08223 AF145042
VERSION	U08223.2 GI:6849140
KEYWORDS	Streptomyces fradiae
SOURCE	Streptomyces fradiae
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 10167)
AUTHORS	Merson-Davies, L.A. and Cundliffe, E.
TITLE	Analysis of five tylosin biosynthetic genes from the tyl1BA region of the Streptomyces fradiae genome
JOURNAL	Mol. Microbiol. 13 (2), 349-355 (1994)
MEDLINE	95075319
PUBMED	7984112
REFERENCE	2 (bases 7082 to 10167)
AUTHORS	Bate, N., Butler, A.R., Gandeche, A.R. and Cundliffe, E.
TITLE	Multiple regulatory genes in the tylosin biosynthetic cluster of Streptomyces fradiae
JOURNAL	Chem. Biol. 6 (9), 617-624 (1999)
MEDLINE	99398833
PUBMED	10467127
REFERENCE	3 (bases 7082 to 10167)
AUTHORS	Bate, N., Butler, A.R., Smith, I.P. and Cundliffe, E.
TITLE	The mycarose-biosynthetic genes of Streptomyces fradiae, producer of tylosin
JOURNAL	Microbiology 146 (Pt 1), 139-146 (2002)
MEDLINE	20121747
PUBMED	10658660
REFERENCE	4 (bases 1 to 10167)
AUTHORS	Merson-Davies, L.A.
TITLE	Direct Submission
JOURNAL	Submitted (31-MAR-1994) Louise A. Merson-Davies, Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK
REFERENCE	5 (bases 1 to 10167)
AUTHORS	Butler, A.R.
TITLE	Direct Submission
JOURNAL	Submitted (23-AUG-1999) Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK
REMARK	Sequence update by submitter
COMMENT	On Feb 1, 2000 this sequence version replaced gi:473596.
FEATURES	Location/Qualifiers

1..10167	/organism="Streptomyces fradiae"
	/mol_type="genomic DNA"
	/strain="T59235"
	/db_xref="taxon:1906"
944..2197	/note="ORF1"
	/codon_start=1
	/transl_table=11
	/product="cytochrome P-450"
	/protein_id="AAA21341.1"
	/db_xref="GI:473597"
translation="MTQTLEAEKPEPERTANSVHPLAQAPAGAGAGLLEWFAFARA EAPVWDESRQAVFYDDYITVTNPLQFSDFSPVPFVPEETAALMGPGTGGID PRHGLRKLVSQATPRRIATLEPRIATRLGLDGLREKQIDVVSIDLALPLVIV IAELLGPAEDRLFRWDVILNNEGMEYPLNLPDFSETMGPAIKWGDYLYRRLAL KRETPDDMLMSGLIENVEGRRLTDEIYVIVALLTAGHISATLGLNLFVLDRHR NQAELRADRLPGAIETRLRSPFNIFRLKEDDTDLGHPMKAGQMVVAVIASA EDHEVEFYBEDELTAARLPTVTRH"	
2677..3843	/note="ORF2. GTG start codon"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAA21342.1"
	/db_xref="GI:473598"
translation="MTGLRPAVRVPHDLROVHAATGVSEIGGALLRVAARGRYLL	



CDS  
GAELAAFEERFAEYCGNAHCVAVGSGLDDARLALMALCVGSGDEVIIVPSHTFIASWLA  
VSATGATFVPEPGSPGGAFLDPDRLEAALTPTTRAVMPVHLVGHVDPDVPVG  
AFAPHLAVDAEAQATARYRGRRIIGSGHRTAFSPYKNGKALGDGGAVTSDPEL  
ADRLRLNVCAREKYRHEERGTSRLDELQAAVLVSKLPLYLDAWNTRRRIIAARYGE  
ALAGLCVTPVPEGVAVPVMHOYVLRSPYRDLRRRLAAGVETLVHPVPAVAVHAGAY  
AGAGCPAGGLPRARLAGEVLSPIGPHLZDEAVEVIAAVQSAALDSWEEGP"  
3840..4751  
/note="ORF3; GTG start codon"  
/codon\_start=1  
/transl\_table=1  
/product="dtdp-glucose synthase"  
/protein\_id="AA021343.1"  
/db\_xref="GI:473599"  
/translation="MNDRPRAMKGIILAGSGTTLRLPLTGLSKQLLVVYDKMIYY  
PLSVMLAGIREIGIISKDLDFLSLGGDLGLSISYAEQREPRGIAEAFLLIGA  
RHIGDDAALITGNVDFPGPSSVLGTVARLDGCELFYGVFKDAHRYGVGEIDSGG  
RLLSLEKPRPLPGRHRLVLYTNDVVEIARTISPSARGLELTVDNVKYLEOGRAA  
HGAGAVANWLDMGTHDSLLQAGOVQLLEQFQGERIACIEBIAMRMGFIASAEQYRLG  
CELKSSSYGSIIDVAMGAADSRAQ"  
4812..5813  
/note="ORF4"  
/codon\_start=1  
/transl\_table=1  
/product="dtdp-glucose dehydratase"  
/protein\_id="AA021344.1"  
/db\_xref="GI:473600"  
/translation="MRVLVTGAGFISGHFTGQLTGAYPDIGATRTVVLDKLTVAGN  
PANLEHVAGHPDLFVFRVGHVADHGWRRRLMEGVGLVHFHAEHSDRSIESSEAFVR  
NVEVTRVLQAAVAGVGRFVHISTDVEYGSIAEGSWPEDHPVAPNSPYAATKAASDL  
LALAHRTYGLDVTRTNNYGRQYPEKAVPLFTNLLDGLPVLPGDGNTRML  
HVDDHCRGVALVAGGAPGVYINIGGTTELNTABLTDLKLELGCADRSALRRVDRPG  
HDRYSYDITKIRBELYAPRTGTIEGLAGTVAMVDRNRAWMEPLKRSPPGRELERA"  
5875..6636  
/note="ORF5; GTG start codon"  
/codon\_start=1  
/transl\_table=1  
/product="thioesterase"  
/protein\_id="AA021345.1"  
/db\_xref="GI:473601"  
/translation="MMLRCAVPVPTPAHRLVCFPHAGSGARAYRPFALAAAGVET  
HAVYQGRDRRKEPFAITTELAERVLPELRLLDAPDVPVVALFGHSMGAVVAYET  
ARLHRSGAPRAGLLSGRAPTADRTETAHLGDLRELLAEIRLQGTDPGALADEB  
VLMVLPAIRGDYAAVGRYRHPVGRPCPLTVFTGADPNVTIPEARMRELITGAF  
ALRVFPGGHFYLNDQRAVACVETIELRLHGSKSAH"  
6763..8130  
/gene="tylCVI"  
/note="required for mycarose biosynthesis"  
/function="required for mycarose biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="TylCVI"  
/protein\_id="AAF29379.1"  
/db\_xref="GI:6849141"  
/translation="KTLPGDIRMARAGRGAPVALIPFAMRGWHEFDPANGLRHE  
SGRFFSLEGVRVGRDGRVWGQPIIVQPEVGLIGIVAREIHHVHLFVQAKMEPGN  
INTLQISPTQATSNFTGRGRIGRIFDLFLEFGRARVLVDVLSQAEWFLAKRN  
RMVVEIOPDELETGDFRMLTGOIRALLTADNVMMDARSVLACLPTAGGQVAG  
DDFAALRRSPTRDSAPTLCSAAELGSMFTGVRALEELVORLIPLDVAAHGWHGPD  
EIAHRSHFRVLADITASNREVARWSQPLIQBHPFSLALIIVKRVDPGLHALVQAR  
VDVGHNVAVLPAITVHCPRADHRAEPAPPIYLDVILGASPAQPRYDTVQSEGGFR  
YHARNYAVIVSPFSDAAGTDEDTAGDYAVMTFPGQITELLAHGNVYNVELRTLIAC  
AHALY"  
complement(8270..9562)  
/gene="tylR"  
complement(8270..9562)  
/gene="tylR"  
/function="regulation of tylosin production"  
/note="tyl orf7"  
/codon\_start=1  
/transl\_table=1  
/product="TylR"

gene  
CDS  
/protein\_id="AAF29380.1"  
/db\_xref="GI:6849142"  
/translation="MSAVTGAATPVTEPWPVLPGRASVPAVRTPDAAAGPCAAAGDICTD  
VLATLRSDQRRKERYVHGLLTHGRITIRNIAAMIHEHAGEQSLHEHFISSSTWMS  
LLRLARLRLBOELASVGVVRRYVPMVVRVPGVGVRRYVPHLQTVNSHOSWGIWY  
LRESGAVPVNWLSTIGDGLMDEGLRRRAAIPRELRARPSEAAVAGIIVSETAGWGLPR  
KPLVWDAELVASLIRALSAGOPFMLRIDNGTLLAPGLSGAGRPVTTAARTAEEL  
ARSQRPVWVDPAEPVPTSLALLPLFCNFGLLPVFGTVARPGAGAGNGTFRPAPR  
SLVLAQVOPDRSRVVELVNMNDAGRTLLRLGLKFLRVTETGAGAGHDVGLRDFE  
GRSYPGWRHVRVTLASLAHAMVLAAPDIPAAAGGRATA"

Query Match 13.1%; Score 156.4; DB 1; Length 10167;  
Best Local Similarity 49.3%; Pred. No. 1.4e-33;  
Matches 409; Conservative 0; Mismatches 421; Indels 0; Gaps 0;  
QY 250 ATGACCCCGCGAAGCATACAAAATCCGTTCCAGTCGTGGAACAAAGCCCTTACTCCGCGC 309  
DB 1244 ATCGACCCCGCCCGCAGCGCCGCTGCGAAGCTGGTCTCGCAGCGGTTCACCCCGCGG 1303  
QY 310 GTGATGAAGCAATGGGAACCGGAGAAATCAAGAAATCAACAGATGAAGTCAAAAATTT 369  
DB 1304 CGGATCGGCACACTGGAACCGCGGATCGCGAGATCACCGCGGATTCGTGGAACGGGTG 1363  
QY 370 CAGGGCGCAGTGTGACCTTGTTCACGATTTTTCATACCCGCTCCGCTTATTGTG 429  
DB 1364 CGGGAAGGGCCAGATCGAGCTCTCCGATCTGGCCTTCCGCTCGCGTGTATCGTC 1423  
QY 430 ATATCTGAGCTGCTGGGAGTGCCTTCAGCGCAGATGGAAAGTCTTAAAGCATGGTCTGAT 489  
DB 1424 ATCCCGAGCTGCTCGGCATACCGCGGAGGACCGGTGATCTCTTCGCGAGTGGGTGCGAC 1483  
QY 490 CTTCTGTCAGTACCGAAGATAAAAGTGAAGAGCTGAAAAGCCCTTTTGGAGAA 549  
DB 1484 GTCATCTCAACACGAGGGCATGATATACCCCAATCTCCCGAGGAGACTTCTCCGAGCG 1543  
QY 550 CGAGATAGTGTGAGGAAGACTGCGCGCTTTTTCGCGGCATCATAGAAAGAAAGCA 609  
DB 1544 ATGGCCCGCCCATCAAGAGTGGGGCGACTACCTGTACCGAGCATCGCGTCAAGCGG 1603  
QY 610 AACAAACCGGAACAGSATAATTTTCTATTTTAGTGAAGCGGAAGAAACAGCGGAGAG 669  
DB 1604 GAGAGCCGAGCGGACGATCTGATGAGCGGGCTCATCGAGGCGGAGTTCGAGGGCGCAGG 1663  
QY 670 CTGTCGCTGAAGAGCTGATTCGGTTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729  
DB 1664 CTCACCGAGGAGGATCGTCAACATCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1723  
QY 730 ACTACAAACCTGATTTCAAAATCGCATGTACACATATTAGAAACGCGAGCGCTTTACGAG 789  
DB 1724 AGCGCCACGCTGCTCGGCAACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1783  
QY 790 GAACTGCGCAGCATCTCTGAATGATGCTTCAGGCAAGTGGAGGAAAGCCCTTGGCTTTTCA 849  
DB 1784 GAGCTGCTGCGCAGCGGTGACCTGATCCCGCGCGCATCGAGGAGAGCGCTGCGCTACCGC 1843  
QY 850 GCGCGCGCCCGGTTTTCAGCGGCAATTCGCAAGCGGGAATCGAGATCGGGGGGACCTG 909  
DB 1844 AGCCCGGTTCACAAACATCTTCGCGCTGCTGGAAGGAGGACACGAGCATCTCCGCCATCCG 1903  
QY 910 ATTAAGAGGAGTATATGTTTGGCGTTTGTGGCATCGGCAATCGTATGAAGCAAG 969  
DB 1904 ATGAAGGCGCGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1963  
QY 970 TTGTGACAGACCGCATGTTTGTATATCCCGCCCATCCCAATCCCGCATATTCGGTTTGGC 1029  
DB 1964 TTCTCGATCCGACACAGTTTCAGCTGCGCGCGGCGGACCAACAGACACATGCTGTTCCGG 2023  
QY 1030 CACGCAATCCATTTTTCGCTGCGGCGCGCGCTTTCGCGCTTTCGAGCAAA 1079  
DB 2024 CACGCGATCCACCACTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2073

RESULT 15

MLPRTN8/3  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

342300 bp DNA linear BCT 22-FEB-2001  
Mycobacterium leprae strain TN complete genome; segm:ent 8/10.  
AL583924 AL450380  
AL583924.1 GI:13093618

Mycobacterium leprae  
Mycobacterium leprae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
1 (bases 1 to 342300)  
Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,  
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,  
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,  
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,  
Hamlin, N., Holroyd, S., Hornby, T., Jagels, K., Lacroix, C.,  
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,  
Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,  
Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,  
Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G.  
Massive gene decay in the leprosy bacillus  
Nature 409 (6823), 1007-1011 (2001)  
2128732  
11234002  
2 (bases 1 to 342300)  
Parkhill, J.  
Direct Submission  
Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium  
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome  
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur 28 rue du Docteur Roux,  
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk

Details of M. leprae sequencing at the Sanger Centre are available  
from [http://www.sanger.ac.uk/projects/M\\_leprae/](http://www.sanger.ac.uk/projects/M_leprae/) A relational  
database containing the M. leprae sequences is available from  
<http://genolist.pasteur.fr/Leprona/>.

Location/Qualifiers  
1 . 342300  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/strain="TN"  
/db\_xref="taxon:1769"  
275..430  
/gene="ML1963"  
/pseudo  
275..430  
/gene="ML1963"  
note="Possible pseudogene of M. tuberculosis orthologue  
Rv3463 (Best blastx score 121)"  
/pseudo  
/codon\_start=1  
/transl\_table=1  
/product="probable neuraminidase (pseudogene)"  
483..1484  
/gene="rmlB"  
/note="synonym: ML1964"  
483..1484  
/gene="rmlB"  
note="Similar to M. tuberculosis dTDP-glucose  
4,6-dehydratase rmlB Rv3464 TR\_006329 (EMBL:Z95390) (331  
aa); Fasta score E(): 0, 84.0% identity in 331 aa overlap,  
and TR:Q50556 (EMBL:U43540) (329 aa); Fasta score E(): 0,  
78.2% identity in 331 aa overlap, and to others e.g.  
Streptococcus pneumoniae dTDP-glucose-4,6-dehydratase cpnB  
TR:Q54611 (EMBL:AF030364) (349 aa); Fasta score E(): 0,  
59.9% identity in 334 aa overlap. Contains Pfam match to  
entry PF01370 Epimerase, NAD dependent  
epimerase/dehydratase family.  
Similar to ML0204, ML0751, ML1942 and ML2428"  
/codon\_start=1  
/transl\_table=1

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2139..2624  
/gene="lpqH"  
note="Similar to M. tuberculosis 19 kDa lipoprotein  
antigen precursor lpqH Rv3763. Contains a possible  
N-terminal signal sequence and a PS00013 Prokaryotic  
membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=1  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/db\_xref="SPTREML:Q9X7A5"  
/translation="MRHKLAAIYAVITMAGACSGGTQAPFVSKTNSSTTTVA  
SSIPDAAGETKVTIGQPQKVGVPVVCSTTNGKFSIAIGMTIGVIGLEPDASVYVH  
NAGLGTIDGVIAITEGVPSNANATNGYTIQITGTASGVDNTGQIHKHSFEIVTC  
R"

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2139..2624  
/gene="lpqH"  
note="Similar to M. tuberculosis 19 kDa lipoprotein  
antigen precursor lpqH Rv3763. Contains a possible  
N-terminal signal sequence and a PS00013 Prokaryotic  
membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=1  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/db\_xref="SPTREML:Q9X7A5"  
/translation="MRHKLAAIYAVITMAGACSGGTQAPFVSKTNSSTTTVA  
SSIPDAAGETKVTIGQPQKVGVPVVCSTTNGKFSIAIGMTIGVIGLEPDASVYVH  
NAGLGTIDGVIAITEGVPSNANATNGYTIQITGTASGVDNTGQIHKHSFEIVTC  
R"

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2139..2624  
/gene="lpqH"  
note="Similar to M. tuberculosis 19 kDa lipoprotein  
antigen precursor lpqH Rv3763. Contains a possible  
N-terminal signal sequence and a PS00013 Prokaryotic  
membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=1  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/db\_xref="SPTREML:Q9X7A5"  
/translation="MRHKLAAIYAVITMAGACSGGTQAPFVSKTNSSTTTVA  
SSIPDAAGETKVTIGQPQKVGVPVVCSTTNGKFSIAIGMTIGVIGLEPDASVYVH  
NAGLGTIDGVIAITEGVPSNANATNGYTIQITGTASGVDNTGQIHKHSFEIVTC  
R"

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2139..2624  
/gene="lpqH"  
note="Similar to M. tuberculosis 19 kDa lipoprotein  
antigen precursor lpqH Rv3763. Contains a possible  
N-terminal signal sequence and a PS00013 Prokaryotic  
membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=1  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/db\_xref="SPTREML:Q9X7A5"  
/translation="MRHKLAAIYAVITMAGACSGGTQAPFVSKTNSSTTTVA  
SSIPDAAGETKVTIGQPQKVGVPVVCSTTNGKFSIAIGMTIGVIGLEPDASVYVH  
NAGLGTIDGVIAITEGVPSNANATNGYTIQITGTASGVDNTGQIHKHSFEIVTC  
R"

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2139..2624  
/gene="lpqH"  
note="Similar to M. tuberculosis 19 kDa lipoprotein  
antigen precursor lpqH Rv3763. Contains a possible  
N-terminal signal sequence and a PS00013 Prokaryotic  
membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=1  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/db\_xref="SPTREML:Q9X7A5"  
/translation="MRHKLAAIYAVITMAGACSGGTQAPFVSKTNSSTTTVA  
SSIPDAAGETKVTIGQPQKVGVPVVCSTTNGKFSIAIGMTIGVIGLEPDASVYVH  
NAGLGTIDGVIAITEGVPSNANATNGYTIQITGTASGVDNTGQIHKHSFEIVTC  
R"

misc\_feature  
2169..2201  
/gene="lpqH"  
note="PS00013 Prokaryotic me-brane lipoprotein lipid  
attachment site"  
complement(4768..4995;  
/gene="ML1967"  
/pseudo  
complement(4769..4995)  
/gene="ML1967"

gene  
CDS

misc\_feature  
1492..2025  
/gene="rmlC"  
note="Pfam match to entry PF00908 dTDP sugar isom",  
dTDP-4-dehydrorhamnose 3,5-epimerase, score 308.80,  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="synonym: ML1966"  
2

```

/note="Possible pseudogene of M. tuberculosis paralog PPE"
/pseudo
/codon_start=1
/transl_table=11
/product="PPE family protein (pseudogene)"
complement(5464..5649)
/gene="ML1968"
gene

/pseudo
complement(5464..5649)
/gene="ML1968"
CDS

/note="Possible pseudogene of M. tuberculosis paralog PE"
/pseudo
/codon_start=1
/transl_table=11
/product="PE family protein (pseudogene)"
complement(5661..8043)
/note="Dispersed repeat, LEPRP, copy 6"
complement(5848..5946)
/gene="ML1969"
gene

/pseudo
complement(5848..5946)
/gene="ML1969"
CDS

/note="Similar to Agrobacterium tumefaciens transposase
TR:Q44454 (EMBL:Z18270) (366 aa) fasta scores: E():
0.0002, 51.5% id in 33 aa, and to Pseudomonas putida
transposase tnpal TR:Q9R9U9 (EMBL:AJ245436) (355 aa) fasta
scores: E(): 0.00088, 48.5% id in 33 aa"
/pseudo
/codon_start=1
/transl_table=11
complement(6211..7354)
/gene="ML1970"
gene

/pseudo
complement(6211..7354)
/gene="ML1970"
CDS

/note="Similar to many e.g. Cryphonectria parasitica
(Chestnut blight fungus) putative maturase TR:AF27656
(EMBL:AF218567) (778 aa) fasta scores: E(): 7.8e-11, 29.3%
id in 216 aa"
/pseudo
/codon_start=1
/transl_table=11
/product="putative group II intron maturase-related
protein"
complement(6674..6727)
/gene="ML1970"
gene

/pseudo
complement(7559..7698)
/gene="ML1971"
CDS

complement(7559..7698)
/gene="ML1971"
CDS

Query Match 12.6%; Score 150; DB 1; Length 342300;
Best Local Similarity 48.8%; Pred. No. 2.1e-31;
Matches 439; Conservative 0; Mismatches 455; Indels 6; Gaps 1;

QY 238 TCATCATTAACATGACCCCGGAGCATACAAATCCGTTTCAGTCGTGAAACAGCC 297
DB 134656 TCGTTTATGTTCTCGATCCACCATATACCCGGTTACGAAAGCTTGTGAGAAAGCG 134597

QY 298 TTACTCCGCGGTGATGAACATGGAGCGGAGMATTCAGAAATCACAGATGAAC 357
DB 134596 TTCGCACGAAAGTCGTTTCAGCGCGCTGGAGGTGATATTGTCGCTGGTGGATTCCCTG 134537

QY 358 ATTCAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCACGATTTTTCATACCGGCT 417
DB 134536 CTGGATAAGGGTGGCGCGGCTGGCCAGATTTCAGTCATTTGCTGACCTAGCCTTCGGCTT 134477

```

```

QY 418 CCGGTTATTGTGATATCTGAGCTGCTGAGTGCCCTTCAGCGCAGATGGAACAGTTTAAA 477
DB 134476 GCGGTGGCTGTGATTTGTTCGGTCTGCTGGGTGTCCGTACGAGACGCGCCGAGTTCGGT 134417

QY 478 GCATGGTCTGATCTTCTGGTCACTACA-----CCGAAGGATAAAAGTGAAGAAAGCTGAA 531
DB 134416 CCGCTCTCAGCACATATTGGTGCAGTCAGTCAGCCGTTTCATTACCATTTACCGGTGAACCG 134357

QY 532 AAAGCCTTTTGGGAACCGAGATAAGTGTGAGAAAGAACTGGCGCGGTTTTTTTGGCCGC 591
DB 134356 CCGGAGGCCCATCTGAAGAGCGGCTCGCGGCTGGTGTGGTTTGGCGGATTACTTGGAGCAG 134297

QY 592 ATCATAGAGAAAGCGAAACAAACCGGAAACAGGATATTATTTCTATTATTTAGTGAAGCG 651
DB 134296 CTAGTTAAATGTCCACGCGGTACACCGGTGAGGATCTGATATCCAGGCTGATTGAGTTG 134237

QY 652 GAAGAAACAGGCGAGAAAGCTGTCCGGTGAAGAGCTGATTCGCTTTTTCGACGCTGCTGCTG 711
DB 134236 GACGAGTCCGGTGATCAGCTGACCGGAGGAGAGATCATCGCAACGTCGGCTCTGTTGCTC 134177

QY 712 GTGCCCGGAAATGAACCACTACAAACCTGATTTCAAAATGCGATGTACAGCATATAGAA 771
DB 134176 GTTCTGCTCATGAGACCAACCGTGAATCTGATTCGCAATGCGGCTCCTGCGGATGTACGC 134117

QY 772 ACGCCAGCGCTTTACGAGGAAGTGGGAGCCATCCTGAACCTGATGCTCAGGCGAGTGGAG 831
DB 134116 AACCCGTGCGAGTGGAAAGCGCTAAGTAGCAACCCGAGCGGCGCGCTGGTGGTTGAG 134057

QY 832 GAAGCCTTTCGCTTTCAGAGCGCGCGCGCTTTTGGAGCGCATTTGCCAAGCGGATAGC 891
DB 134056 GAGACGCTTCGTTATGATCCAGCAATCCATCTAATCGGTGCGGTAGCTGCTAAGGACATG 133997

QY 892 GAGATCGGGGGCGACCTGATTTAAAGAGGTGATATGGTTTTGGCGTTTGTGGCATCGCA 951
DB 133996 ACGATCGGTCAAGCAACCTTTAAACGAGGCTGATACAATGCTTTTGTGCTTGGCGCGCC 133937

QY 952 AATCGTGAATGAAGCAAGTTTGGACAGCGGCACATGTTTGATATCCGCGCGCATCCCAAT 1011
DB 133936 AATCGGATCCAGCGGTTTACTCAGCCCCCGATGATNTTGGACCCGACCGTCCAGCTCA 133877

QY 1012 CCGCATATTGGTTTGGCCACGCGCATCCATTTTTCCTTGGGGCGCCGCTTGGCCGCTTT 1071
DB 133876 CGACATTTGGCAATTTGCGTTCGGCTCAATTTTGGCTAGAGGCGCATTTGCTCGGTTG 133817

QY 1072 GAAGCAATATCGGCTTAACGCTCTTTGTTGCTTTTCTCTCATATGAGAGTGGCTGAGT 1131
DB 133816 GAAGCGACTGTCAATATCTCGATCAGTCGCGGCTTTCCACAGGTACAGCTGCCCGGT 133757

```

Search completed: May 28, 2004, 21:28:05  
Job time : 3283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:08:41 ; Search time 223 Seconds  
(without alignments)  
501.744 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052  
Sequence: 1 MNVLNRQALQALLNGKK.....SITPIENSVIYGLKSRVVM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2052	100.0	396	3 AAB15501	Aab15501 Bacillus
2	2042	99.5	396	3 AAB15502	Aab15502 Bacillus
3	2018	98.3	396	3 AAB15503	Aab15503 Bacillus
4	670	32.7	410	6 AAE36997	Aae36997 Micromono
5	649.5	31.7	404	6 ABR82138	AbR82138 Bacterial
6	642	31.3	416	3 AAB18654	Aab18654 Amino aci
7	642	31.3	416	3 AAY67218	Aay67218 Macrolide
8	642	31.3	416	6 ABG71678	Abg71678 S. venezu
9	642	31.3	416	6 ADA09417	Ada09417 S. venezu
10	636	31.0	425	6 ABR82128	AbR82128 Bacterial
11	629.5	30.7	415	3 AAY77196	Aay77196 S. venezu
12	626	30.5	399	6 ABP57702	Abp57702 Saccharop
13	623.5	30.4	392	6 ABR82134	AbR82134 Bacterial
14	623	30.4	396	5 ABB81315	Abb81315 Streptomy
15	606.5	29.6	398	2 AAW11585	Aaw11585 Streptomy
16	587.5	28.6	408	6 ABR82120	AbR82120 Bacterial
17	586	28.6	404	6 ABR82143	AbR82143 Bacterial
18	581	28.3	396	2 AAR60777	Aar60777 Mycinamic
19	576.5	28.1	428	7 ADD13339	Add13339 C. glutam
20	576.5	28.1	430	4 AAG90363	Aag90363 C. glutam
21	569.5	27.8	428	4 AAB79632	Aab79632 Corynebac
22	565	27.5	399	6 ABU11376	Abu11376 Protein e
23	559	27.2	403	7 ADD15221	Add15221 EryF prot
24	550	26.8	432	6 AAE34726	Aae34726 Streptomy
25	549	26.8	429	6 AAE34728	Aae34728 Streptomy

26	549	26.8	430	6 AAE34780	Aae34780 Streptomy
27	549	26.8	430	6 AAE34736	Aae34736 Streptomy
28	547	26.7	430	6 AAE34725	Aae34725 Streptomy
29	547	26.7	430	6 AAE34731	Aae34731 Streptomy
30	544	26.5	404	2 AAR14724	Aar14724 6-hydroxy
31	544	26.5	430	6 AAE34727	Aae34727 Streptomy
32	540	26.3	430	6 AAE34734	Aae34734 Streptomy
33	539	26.3	415	6 ABR82141	AbR82141 Bacterial
34	537.5	26.2	430	6 AAE34735	Aae34735 Streptomy
35	537.5	26.2	430	6 AAE34723	Aae34723 Streptomy
36	536	26.1	430	6 AAE34730	Aae34730 Streptomy
37	532.5	26.0	411	2 AAW54389	Aaw54389 Actinomad
38	531	25.9	430	6 AAE34733	Aae34733 Streptomy
39	531	25.9	430	6 AAE34737	Aae34737 Streptomy
40	530	25.8	426	6 AAE34738	Aae34738 Streptomy
41	530	25.8	430	6 AAE34722	Aae34722 Streptomy
42	530	25.8	475	6 AAE34732	Aae34732 Streptomy
43	526	25.6	482	5 ABB81316	Abb81316 Streptomy
44	523	25.5	415	6 ABR82146	AbR82146 Bacterial
45	521	25.4	475	6 AAE34729	Aae34729 Streptomy

#### ALIGNMENTS

RESULT 1  
AAB15501  
ID AAB15501 standard; protein; 396 AA.  
XX  
AC AAB15501;  
XX  
DT 14-FEB-2001 (first entry)  
XX  
DE Bacillus subtilis hydroxylating protein.  
XX  
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;  
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.  
XX  
OS Bacillus subtilis.  
XX  
PN W0200044886-A1.  
XX  
PD 03-AUG-2000.  
XX  
PF 28-JAN-2000; 2000WO-JP0000472.  
XX  
PR 29-JAN-1999; 99JP-00021707.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX  
WI: 2000-548827/50.  
DR N-PSDB; AAA95666.  
XX  
PT New protein derived from Bacillus genus microorganism useful for  
producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
inhibitors.  
XX  
PS Claim 8; Page 68-71; 111pp; Japanese.  
CC This sequence represents a novel Bacillus derived protein having the  
activity of producing a hydroxylated bicyclic compound or the  
corresponding lactone from a bicyclic compound or the corresponding  
lactone. The protein is used for preparing hydroxylated compounds useful  
as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and  
hypocholesterolaemic agents  
XX  
SQ Sequence 396 AA;

Query Match 100.0%; Score 2052; DB 1; Length 396;  
Best Local Similarity 100.0%; Pred. No. 7.2e-180; Indels 0; Gaps 0;  
Matches 396; Conservative 0; Mismatches 0;

QY 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60  
 DB 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60  
 QY 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKQWEPRIOEITDELI 120  
 DB 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKQWEPRIOEITDELI 120  
 QY 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMQFKAWSDLVSTPKDKSEAEKAF 180  
 DB 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMQFKAWSDLVSTPKDKSEAEKAF 180  
 QY 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240  
 DB 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240  
 QY 241 NETTTLISNAMYSIILETPGVYEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300  
 DB 241 NETTTLISNAMYSIILETPGVYEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300  
 QY 301 GHLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFHFCIGAPLARLEAN 360  
 DB 301 GHLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFHFCIGAPLARLEAN 360  
 QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396  
 DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396

## RESULT 2

AAB15502  
 ID AAB15502 standard; protein; 396 AA.  
 AC AAB15502;  
 DT 14-FEB-2001 (first entry)  
 DE Bacillus subtilis hydroxylating protein #2.  
 KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;  
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.  
 XX Bacillus subtilis.  
 OS Bacillus subtilis.  
 XX WO200044886-A1.  
 PN 03-AUG-2000.  
 XX 28-JAN-2000; 2000WO-JP000472.  
 XX 29-JAN-1999; 99JP-00021707.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
 XX WPI; 2000-548827/50.  
 DR N-PSDB; AAA95667.  
 XX New protein derived from Bacillus genus microorganism useful for  
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
 PT inhibitors.  
 XX Claim 10; Page 94-96; 11pp; Japanese.  
 PS This sequence represents a novel Bacillus derived protein having the  
 CC activity of producing a hydroxylated bicyclic compound or the  
 CC corresponding lactone from a bicyclic compound or the corresponding  
 CC lactone. The protein is used for preparing hydroxylated compounds useful  
 CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and  
 CC hypocholesterolaemic agents

## SQ Sequence 396 AA;

Query Match 99.5%; Score 2042; DB 3; Length 396;  
 Best Local Similarity 99.5%; Pred. No. 6e-179;  
 Matches 394; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60  
 DB 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60  
 QY 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKQWEPRIOEITDELI 120  
 DB 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKQWEPRIOEITDELI 120  
 QY 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMQFKAWSDLVSTPKDKSEAEKAF 180  
 DB 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMQFKAWSDLVSTPKDKSEAEKAF 180  
 QY 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240  
 DB 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240  
 QY 241 NETTTLISNAMYSIILETPGVYEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300  
 DB 241 NETTTLISNAMYSIILETPGVYEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300  
 QY 301 GHLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFHFCIGAPLARLEAN 360  
 DB 301 GHLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFHFCIGAPLARLEAN 360  
 QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396  
 DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396

## RESULT 3

AAB15503  
 ID AAB15503 standard; protein; 396 AA.  
 AC AAB15503;  
 XX 14-FEB-2001 (first entry)  
 DE Bacillus subtilis hydroxylating protein #3.  
 XX Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;  
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.  
 XX Bacillus subtilis.  
 OS Bacillus subtilis.  
 XX WO200044886-A1.  
 PN 03-AUG-2000.  
 XX 28-JAN-2000; 2000WO-JP000472.  
 XX 29-JAN-1999; 99JP-00021707.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
 XX WPI; 2000-548827/50.  
 DR N-PSDB; AAA95669.  
 XX New protein derived from Bacillus genus microorganism useful for  
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
 PT inhibitors.  
 XX Claim 10; Page 105-107; 11pp; Japanese.  
 PS This sequence represents a novel Bacillus derived protein having the  
 CC activity of producing a hydroxylated bicyclic compound or the

CC corresponding lactone from a bicyclic compound or the corresponding  
CC lactone. The protein is used for preparing hydroxylated compounds useful  
CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and  
CC hypocholesterolaemic agents  
XX  
SQ Sequence 396 AA;

Query Match 98.3%; Score 2018; DB 3; Length 396;  
Best Local Similarity 98.2%; Pred. No. 9.6e-177;  
Matches 389; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNVLRNRQALQALINGKQKQDAYHPFPWYESMRKQAPVSPDEENQVMSVLYDDVKVV 60  
DB 1 MNVLRNRQALPRALLNGKQKQDAYHPFPWYESMRKQAPVSPDEENQVMSVLYDDVKVV 60  
QY 61 GQKELFSSCMPQOTSIGNSIINMPDPKTKRSVVKAFPTPRVWKQWEPRIQEITDELI 120  
DB 61 GQKELFSSCMPQOTSIGNSIINMPDPKTKRSVVKAFPTPRVWKQWEPRIQEITDELI 120  
QY 121 QKFGQSEFDLVHDFSYPLVIVISSELLGVPSAHMEQFKAWSDLVSTTPKDKSEAEKAF 180  
DB 121 QKFGQSEFDLVHDFSYPLVIVISSELLGVPSAHMEQFKAWSDLVSTTPKDKSEAEKAF 180  
QY 181 LEERDKCEELAAFPAGIIEERKNKPEQDIISILVEABETGKLSGEELIPCTLLVAG 240  
DB 181 LEERDKCEELAAFPAGIIEERKNKPEQDIISILVEABETGKLSGEELIPCTLLVAG 240  
QY 241 NETTNTNISNAMYSILETQGVVEELSHPELMPQAVEALRFPAPVLRRIAKRDEIG 300  
DB 241 NETTNTNISNAMYSILETQGVVEELSHPELMPQAVEALRFPAPVLRRIAKRDEIG 300  
QY 301 GHLIKEGQVLAFAVASANRDEAKFDRPHMFDIRRHENPHIAFGHGHFCLGAPLARLEAN 360  
DB 301 GHLIKEGQVLAFAVASANRDEAKFDRPHMFDIRRHENPHIAFGHGHFCLGAPLARLEAN 360  
QY 361 IALTSLISAFPHMECVSIPIENSVIYGLKSFVKM 396  
DB 361 IALTSLISAFPHMECVSIPIENSVIYGLKSFVKM 396

RESULT 4  
AAE36997  
ID AAE36997 standard; protein; 410 AA.  
XX  
AC AAE36997;  
XX  
DT 23-OCT-2003 (revised)  
DT 07-AUG-2003 (first entry)  
XX  
DE Micromonospora carbonacea polyketide synthase (PKS) type I #3.  
XX  
KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme.  
XX  
CS Micromonospora carbonacea.  
XX  
PN CA2391131-A1.  
XX  
PD 19-NOV-2002.  
XX  
PF 26-JUL-2002; 2002CA-07391131.  
XX  
PR 26-JUL-2001; 2001US-0307629P.  
XX  
PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
PI Yang X, Staffa A, Farnet CM;  
XX  
DR WPI; 2003-343556/33.  
DR N-PSDB; AAD55813.  
XX  
PT Novel isolated polypeptide involved in biosynthesis of macrolides by  
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,  
PT preferably for biosynthesis of rosaramicin.

XX  
PS Claim 13; Page 99-101; 206pp; English.  
XX  
CC The invention relates to genes and proteins involved in the biosynthesis  
CC of macrolides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for rosaramicin (a 16-member  
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macrolides by microorganisms. It allows  
CC direct manipulation of the proteins involved in the biosynthesis of  
CC chemical engineering of the proteins involved in the biosynthesis of  
CC rosaramicin. It is useful to catalyse certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polyketide, polyketide substrate or its precursor. The present sequence  
CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23  
CC -OCT-2003 to standardise OS field)

XX  
SQ Sequence 410 AA;  
Query Match 32.7%; Score 670; DB 6; Length 410;  
Best Local Similarity 40.1%; Pred. No. 1e-52;  
Matches 143; Conservative 65; Mismatches 141; Indels 8; Gaps 2;

QY 22 DAYHPPFPWYESMRKQAPVSPDEENQVMSVLYDDVKVVGDKELFSS-----CXPQOTS 75  
DB 19 DAQQLDWFAYMKNWPNVSWDETRQAHVFSYRDYQTVTTNPLIFSSDFTSVFPVPSLA 78  
QY 76 SI--GNSIINMPDPKTKRSVVKAFPTPRVWKQWEPRIQEITDELIQFGSSEFDLVH 133  
DB 79 LLMPGPTIGGIDEPHAPKLVSQAPTPRRIAQMLRIGQITADVLDQVQVDDQRIIDIAS 138  
QY 134 DFSYPLPVIIVISSELLGVPSAHMEQFKAWSDLVSTTPKDKSEAEKAFLEERDKCEELAA 193  
DB 139 DLAYPLPVTIVIAELGPTKDHKEFRWVDIILSNEGLEVPNLDDTETVGPALIEEMSE 198  
QY 194 FFAGIIIEERKNKPEQDIISILVEABETGKLSGEELIPCTLLVAGNETTNTNISAMY 253  
DB 199 FLXQAIAHKAEPKDKLISGLCAAEVDGRKLTDEEVVNIIVALLTAGHISATLESKLEL 258  
QY 254 SILETQGVVEELSHPELMPQAVEALRFPAPVLRRIAKRDEIGHGHFCLGAPLARLEAN 313  
DB 259 VLEEHPQAAVADRSLVPGVIEETLURYSPNCIFRIINEDTDILGHMPKQKQVIAW 318  
QY 314 VASANRDEAKFDRPHMFDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSLISAF 370  
DB 319 IASANRDETEVFTDPTDIEDIRRESNKLHAGHGHFCLGAPLARLEANVFLNQTLDQF 375

RESULT 5  
ABR82138  
ID ABR82138 standard; protein; 404 AA.  
XX  
AC ABR82138;  
XX  
DT 30-SEP-2003 (first entry)  
XX  
DE Bacterial P450 enzyme SEQ ID NO:40.  
XX  
KW Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;  
KW diol; alkene; chirality; thermotolerance; thermostability.  
XX  
OS Eubacteria.  
XX  
PN WO2003052050-A2.  
XX  
PD 26-JUN-2003.  
XX  
PF 05-AUG-2002; 2002WO-US024910.  
XX  
PR 03-AUG-2001; 2001US-0309497P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;

XX WPI; 2003-541641/51.  
 DR N-PSDB; ACP06128.  
 XX Novel polypeptide having P450 enzyme activity and polynucleotides  
 PT encoding the polypeptide, useful for catalyzing the hydrolysis of  
 PT epoxides and arene oxides to their corresponding diols.  
 XX Claim 104; Page 73; 365pp; English.  
 XX The present sequence represents a bacterial P450 enzyme. P450 enzymes can  
 CC be used to catalyze the hydrolysis of epoxides and arene oxides to their  
 CC corresponding diols. P450 enzymes can also be used for hydrolysing an  
 CC alkene, for producing a compound of a desired chirality, and for  
 CC increasing thermostolerance or thermostability of a P450 polypeptide  
 XX Sequence 404 AA;  
 SQ  
 Query Match 31.7%; Score 649.5; DB 6; Length 404;  
 Best Local Similarity 38.3%; Pred. No. 7.7e-51;  
 Matches 143; Conservative 69; Mismatches 122; Indels 39; Gaps 8;  
 QY 26 PFPWYEMKRDAPVSFDEENQ-----VMSVFLYDDVKKVGDKELEPSSC-- 69  
 DB 18 FYPNEADGISLADAYBEAREQGLLRVRMAYGBPAWLATRYADARLVLDGR-PSRAEG 76  
 QY 70 ----MPQOTSSIGNS-IINMDPPKHTKIRSVVNKAFTRPVKQWEPRIQETIDELIQKFQ 124  
 DB 77 ARHDEPQSEGRSGLSDMDPDHTRTLTLVAKFTMHQVEKLRPAVRELADLIDRW 136  
 QY 125 GR-SEFLVHDVSPVPLVIVISELLGVPSAHMOPKAWSDLLVSTPKDKSEAEKAFLE 183  
 DB 137 ATGAPDLVEEFALPVPVGVICQLLGVPEVDRPFRAMSDAALST-----SSLTAE 188  
 QY 184 RDCBEEELAAFFAGIEEKRNKPDQIISILVEAETGEKLSGSELIPCTLLVAGNET 243  
 DB 189 PDANQEEELRAYMGLIEDHRAARPREDLTIGLEARDRDRUTEQELVDLCVGLVAGHET 248  
 QY 244 TTNLISNAMYSLTEPGVYELRSHPELMPQAVREALFRAPAV-----LRRIAKRDT 298  
 DB 249 TATQIPNFVTLDRPEQWNLREDPELVPTAVBELMF---VPLGSGASFPRVATEDVE 305  
 QY 299 IGGHLIKEGOMVLAFVASANDEAKFDRPHMFDIRRHNPNIHAFGHGHCIGAPLARLE 358  
 DB 306 VGGTLVRAGEPVLVAVGAANDPAPFAPQBLDLAREGQHGLFGHGVHHCIGAPLARLE 365  
 QY 359 ANIALTSLSISAFP 371  
 DB 366 LQELGALLRLRP 378  
 RESULT 6  
 AAB18654  
 ID ABE18654 standard; protein; 416 AA.  
 XX AAB18654;  
 AC AAB18654;  
 XX 22-JAN-2001 (first entry)  
 DT  
 XX Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase.  
 DE  
 XX Naibonolide synthase; polyketide synthase gene; narbonolide polyketide;  
 KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;  
 KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;  
 KW picromycin biosynthesis;  
 KW picromycin/methymycin cytochrome P450 hydroxylase.  
 XX Streptomyces venezuelae.  
 OS  
 XX US6117659-A.  
 FW  
 XX 12-SEP-2000.  
 ED  
 XX

PF 27-MAY-1999; 99US-00320878.  
 XX 30-APR-1997; 97US-00846247.  
 PR 06-MAY-1998; 98US-00073538.  
 PR 28-MAY-1998; 98US-0087080P.  
 PR 28-AUG-1998; 98US-00441908.  
 PR 22-SEP-1998; 98US-0100880P.  
 PR 08-FEB-1999; 99US-0119139P.  
 PR 20-MAY-1999; 99US-0134990P.  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 PA Ashley G, Betlach MC, Betlach M, Tang L, Medaniel R;  
 PI WPI; 2000-610844/58.  
 DR  
 XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful  
 PT for converting ketolides to antibiotics and as antibiotics and  
 PT intermediates in the synthesis of compounds with pharmaceutical value.  
 PS Disclosure; Col 47-48; 117pp; English.  
 XX The present sequence represents a picromycin/methymycin cytochrome P450  
 CC hydroxylase (pick). The nucleotide sequence encoding it is used in the  
 CC course of the invention. The specification describes a recombinant DNA  
 CC compound expressing recombinant polyketide synthase genes in host cells  
 CC for the production of narbonolide, narbonolide derivatives and  
 CC polyketides that are useful as antibiotics and as intermediates in the  
 CC synthesis of compounds with pharmaceutical value. The DNA compounds may  
 CC also encode a C12-hydroxylase (pick), desosamine biosynthesis and  
 CC desosaminyl transferase enzymes (useful for conversion of ketolides to  
 CC antibiotics), and the beta-glucosidase enzyme (involved in picromycin  
 CC biosynthesis). These compounds are also useful for increasing the  
 CC antibiotic activity of a compound relative to the unhydroxylated  
 CC compound. The recombinant host cells are useful as genetic systems that  
 CC allow rapid engineering of the narbonolide polyketide synthase. These  
 CC would be valuable for creating novel ketolide analogs for pharmaceutical  
 CC applications  
 XX Sequence 416 AA;  
 SQ  
 Query Match 31.3%; Score 642; DB 3; Length 416;  
 Best Local Similarity 37.4%; Pred. No. 3.9e-50;  
 Matches 148; Conservative 77; Mismatches 133; Indels 34; Gaps 12;  
 QY 21 QD-AVHPFPWYEMKRDAP---VSPDEENQVMSVFLYDDVKKVGD----KELPSSCMP- 71  
 DB 22 QDFAADPYTYARLRAEGFAHRVTRPESDEVLVVGYDSRAVLADPRFXDWRNSTTPL 81  
 QY 72 -QQTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVKQWEPRIQETIDELIQKF-----QGR 126  
 DB 82 TEAEALNHNMLSEDPPTHTKIRLVAAREFTMRVRELLRPVQEIIVGVLADMLAAPGR 141  
 QY 127 SEFDLVHDVSPVPLVIVISELLGVPSAHMOPKAWSDLLVSTPKDKSEAEKAFLEERDK 186  
 DB 142 A--DLMESLAWPLPTIVISELLGVPEPDRAPRVRTDAFVP--PDPAQAQTAM----- 191  
 QY 187 CEEELAAFFAGIEEKRNKPDQIISILVEAETGEKLSGSELIPCTLLVAGNETTT 245  
 DB 192 --AEMSGVSLRLSDSKRGQDGEOLLALVRSDEDSRLTSEELGMAHILLVAGHETTV 249  
 QY 246 NLISNAMYSLTEPGVYELRSHPELMPQAVREALFRAPA-PVLRRTAKRDTGICGHLI 304  
 DB 250 NLLANGYALLSHPDQALALRADMTLLDGAVEMLRYEGPVESATYRFPVEVDJGTVI 309  
 QY 305 KEGDMVLAFVASANDEAKFDRPHMFDIRRHNPNIHAFGHGHCIGAPLARLEANIAT 364  
 DB 310 PAGDTVLVVLADAHHTPERFPDPHFRDITRDTAGHLAFGHGHCIGAPLARLEARIATV 369  
 QY 365 SILISAPPHMECVSTPIE-----NSVIYGLKSFVK 395  
 DB 370 ALLERCPDL-ALDVSFSGELVWVYFNPMPMIRGLKALPIR 404

Qy	21	OD	AYHPFPAYESMRKADP	---VSFOENQVSVFLVDVKVVG	---KELFSSCMP	71
Th	22	OD	FAADPYPYARLRAGGPAHRV	TPEGDEVLVVGVDRAVLADP	PF SKWENSTTL	81



cell. The recombinant methods and materials are useful for expressing polypeptides with significant antibiotic activity, derived in whole or in part from the narboxonolide PKS gene, and other genes involved in narboxonolide and picromycin biosynthesis in recombinant host cells. The present sequence represents *S. venezuelae* picromycin/methymycin cytochrome P450 hydroxylase, Pick.

XX Sequence 416 AA;

Query Match 31.3%; Score 642; DB 6; Length 416;  
 Best Local Similarity 37.4%; Pred. No. 3.9e-50;  
 Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYSEMRKADP---VSFDEENQVMSVFLYDDVKKVVG---KELFSSCMP- 71  
 DB 22 QDFAADPYTYARLAEAGPAHRVTPGEGVWLVGYDRARAVLADPFSKDWNRSTTEL 81

QY 72 -QOTSSIGNSIINMDPKTKIRSVVNKAPTRVMKQWEPRIQETDELIQKF---QGR 126  
 DB 82 TEAEALNHNMLESDPPRTRLRKLVAREFTMRVVELLPVQEIYDGLVDAMLAPDGR 141

QY 127 SEFDLVHDFSYPLPVTVISELLGVPSAHMEQKAWSDLLVSTPKKSEAEAKFLERDK 186  
 DB 142 A--DLMESLAWPLPTVISELLGVPEPDRAFRVWTDVAFV--PDDPAQAQTAM----- 191

QY 187 CEEELAAFFAGLIEKRNKPEQDIISILVE-AEETGKLSGSELIPFCITLLVAGNETTT 245  
 DB 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEEMLYEGVESATYRFPVVDLGGTVI 309

QY 305 KEGDMVLAFVANSANDEAKFDPHMFDIRRHNPHIAGHGIFLCGAPLARLEANIALT 364  
 DB 310 PAGDTVLVVLADAHRTPEFPDPHFRDIRDTAGHLAFGEGHFCIGAPLARLEARI 369

QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSRVK 395  
 DB 370 ALLERCPLD-ALDVSPGELVWYVNPMPMIRGLKALPIR 404

RESULT 9  
 ADA09417  
 XX ADA09417 standard; protein; 416 AA.  
 AC ADA09417;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE *S. venezuelae* picromycin/methymycin cytochrome P450 hydroxylase (Pick).  
 XX  
 KW Streptomyces venezuelae; Sv; narboxonolide polyketide synthase; PKS;  
 KW narboxonolide PKS; narboxonolide biosynthesis; picromycin biosynthesis;  
 KW PKS gene cluster; picromycin; glycosylation; hydroxylation;  
 KW C12 hydroxylase; Pick; desosamine biosynthesis;  
 KW desosaminyl transferase enzyme; antibiotic; narboxonolide synthase;  
 KW picromycin/methymycin cytochrome P450 hydroxylase; Pick; enzyme.  
 XX  
 OS Streptomyces venezuelae.  
 XX  
 PN US6509455-B1.  
 XX  
 PD 21-JAN-2003.  
 XX  
 PF 07-SEP-2000; 2000US-00657440.  
 XX  
 PR 30-APR-1997; 97US-00846247.  
 PR 06-MAY-1998; 98US-00073538.  
 PR 28-MAY-1998; 98US-0087080P.  
 PR 28-AUG-1998; 98US-00141908.  
 PR 22-SEP-1998; 98US-0100880P.  
 PR 08-FEB-1999; 99US-0119139P.

PR 20-MAY-1999; 99US-0134990P.  
 PR 27-MAY-1999; 99US-00320878.  
 XX  
 PA {KOSA-} KOSAN BIOSCIENCES INC.  
 XX Ashley G, Belach MC, Betlach M, McDaniel R, Tang L;  
 PI WPI; 2003-352291/33.  
 XX  
 DR Novel recombinant DNA compounds comprising coding sequences for  
 PT desosamine transferase gene of Streptomyces venezuelae, useful for  
 PT producing desosamine transferase which transfers desosamine to substrate  
 PT polyketides.  
 XX  
 PS Disclosure; Col 77-78; 132pp; English.  
 XX  
 CC The present invention relates to recombinant DNA compounds that encode  
 CC Streptomyces venezuelae (Sv) narboxonolide polyketide synthases (PKSs). The  
 CC recombinant PKSs are derived from narboxonolide PKS and other genes  
 CC involved in narboxonolide and picromycin biosynthesis in recombinant host  
 CC cells. The invention also discloses the *S. venezuelae* PKS gene cluster  
 CC that results in the production of picromycin. Also disclosed are enzymes  
 CC such as those responsible for glycosylation and hydroxylation, (e.g. C12  
 CC hydroxylase (Pick)), desosamine biosynthesis, and desosaminyl transferase  
 CC enzymes. The recombinant narboxonolide, narboxonolide derivatives, and  
 CC polyketides are useful as antibiotics and as intermediates in the  
 CC synthesis of compounds for pharmaceutical applications. The present  
 CC sequence represents *S. venezuelae* picromycin/methymycin cytochrome P450  
 CC hydroxylase (Pick).  
 XX  
 SQ Sequence 416 AA;

Query Match 31.3%; Score 642; DB 6; Length 416;  
 Best Local Similarity 37.4%; Pred. No. 3.9e-50;  
 Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYSEMRKADP---VSFDEENQVMSVFLYDDVKKVVG---KELFSSCMP- 71  
 DB 22 QDFAADPYTYARLAEAGPAHRVTPGEGVWLVGYDRARAVLADPFSKDWNRSTTEL 81

QY 72 -QOTSSIGNSIINMDPKTKIRSVVNKAPTRVMKQWEPRIQETDELIQKF---QGR 126  
 DB 82 TEAEALNHNMLESDPPRTRLRKLVAREFTMRVVELLPVQEIYDGLVDAMLAPDGR 141

QY 127 SEFDLVHDFSYPLPVTVISELLGVPSAHMEQKAWSDLLVSTPKKSEAEAKFLERDK 186  
 DB 142 A--DLMESLAWPLPTVISELLGVPEPDRAFRVWTDVAFV--PDDPAQAQTAM----- 191

QY 187 CEEELAAFFAGLIEKRNKPEQDIISILVE-AEETGKLSGSELIPFCITLLVAGNETTT 245  
 DB 192 --AEMSGYLRLIDSKRGQDGLLSALVRTSDEGSRUTSELLGMHILLVAGHETTV 249

QY 246 NLIANGMYALLSHPDQALALRADMTLLDGAVEEMLYEGVESATYRFPVVDLGGTVI 304  
 DB 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEEMLYEGVESATYRFPVVDLGGTVI 309

QY 305 KEGDMVLAFVANSANDEAKFDPHMFDIRRHNPHIAGHGIFLCGAPLARLEANIALT 364  
 DB 310 PAGDTVLVVLADAHRTPEFPDPHFRDIRDTAGHLAFGEGHFCIGAPLARLEARI 369

QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSRVK 395  
 DB 370 ALLERCPLD-ALDVSPGELVWYVNPMPMIRGLKALPIR 404

RESULT 10  
 ABR82128  
 ID ABR82128 standard; protein; 425 AA.  
 XX  
 AC ABR82128;  
 XX  
 DT 30-SEP-2003 (first entry)  
 XX

DE Bacterial P450 enzyme SEQ ID NO:20.
XX Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
XX diol; alkene; chirality; thermotolerance; thermostability.
XX Eubacteria.
XX WC2003052050-A2.
XX 26-JUN-2003.
XX 05-AUG-2002; 2002WC-US024910.
XX 03-AUG-2001; 2001US-0309497P.
XX (DIVE-) DIVERSA CORP.
XX Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;
XX WPI; 2003-541641/51.
XX N-PSDB; ACF06118.
XX Novel polypeptide having P450 enzyme activity and polynucleotides
XX encoding the polypeptide, useful for catalyzing the hydrolysis of
XX epoxides and arene oxides to their corresponding diols.
XX Claim 104; Page 60-61; 365pp; English.
XX The present sequence represents a bacterial P450 enzyme. P450 enzymes can
XX be used to catalyse the hydrolysis of epoxides and arene oxides to their
XX corresponding diols. P450 enzymes can also be used for hydrolysing an
XX alkene, for producing a compound of a desired chirality, and for
XX increasing thermotolerance or thermostability of a P450 polypeptide
XX Sequence 425 AA;
XX
XX Query Match 31.0%; Score 636; DB 6; Length 425;
XX Best Local Similarity 33.8%; Pred. No. 1.4e-49;
XX Matches 137; Conservative 79; Mismatches 145; Indels 44; Gaps 8;
XX
XX 23 AYHPFPWESMRKDAPV---SFDEENQVMSVFLYDDVKVVGDKELF----- 66
XX 25 ASDPYPAYAMLREHAPVHRITLPSGVEAMLVTRYGDARQALADQLSKNPAHDESPHAK 84
XX
XX 67 --SSCMEQQTSSTGNSINMDPPKHTKERSVNVKAPTFRVWKQWEPRIQETDELIQKQ 124
XX 85 GKTGIGIFGRKAEMLTHLLNIDPPDHTLRLLVSKAFTPRRVAEFTPRVQELTDLDAFV 144
XX
XX 125 GRSEFDLVHDFSVLPVIVTSELGVPSAHMGEQKAWSDLLV---STPKDSEEAEEKAF 181
XX 145 TKSADLIHDFAPFLPIVAICDLGLVPEEDQDDFDWAGMIRHGGGRGGVARSVK--- 201
XX
XX 182 EERDKCEEELAAPFAGIIEEKNRPEQDIISILVEAETGKLGSEELIPCTLLLVAGN 241
XX 202 -----KMRGYLAELTHRKREAPGDGLISGLIKASDHGHELTENEAAAFAFLLFAGF 253
XX
XX 242 ETTTNLISNAMYSLTEPGVVEELRS-----HPELMPQAVEALFRFAPAVLR-RIAKR 295
XX 254 ETTVNLGNGVYQLRHFGPQGRLOTSAAAGETGLLETGIELLRYDGFVEMATWRYATE 313
XX
XX 296 DTEIGGHILKEGDMVLAFVASANDEAKFDRPHFPDIRRHNPNIHAFGHGHIICLGAFLA 355
XX 314 PLTIGGGQDI PASDPVLVLAANDRDPERFDRPDVLDIARRNQHLGYGHGHIYCLGAFLA 373
XX
XX 356 RLEANIALTSLISAPPHMECVSITPIE-----NSVIYGLKSRVK 395
XX 374 RLEGQTALATLLTRLEDLR-LAADPAELWRGGLIMRGLTLPVE 417
XX
XX RESULT 11
XX AAY77196
XX ID AAY77196 standard; protein; 415 AA.
XX XX

AA777196;

12-SEP-2003 (revised)

05-JUN-2000 (first entry)

S. venezuelae macrolide biosynthetic enzyme PikC, SEQ ID NO:39.

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin; neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase; biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma; chronic obstructive pulmonary disease; respiratory inflammation; hypercholesterolaemia; crop protection agent.

Streptomyces venezuelae; ATCC15439.

Key Location/Qualifiers

Misc-difference 103

/note= "Encoded by CCG"

WO200000620-A2.

06-JAN-2000.

25-JUN-1999; 99MO-US014398.

26-JUN-1998; 98US-00105537.

(MINU ) UNIV MINNESOTA.

Sherman DH, Liu H, Xue Y, Zhao L;

WPI; 2003-160679/14.

N-PSDB; AAZ87301.

Desosamine and macrolide biosynthetic gene clusters, useful for, e.g. synthesis of methymycin and pikromycin.

Claim 19; Page 429-430; 438pp; English.

The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or *Streptomyces antibioticus*. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesize methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. Sequences AA77190-77197 represent macrolide biosynthetic enzymes from *Streptomyces venezuelae* ATCC 15439, which are encoded by sequences AAZ87295-287302. (Updated on 12-SEP-2003 to standardise OS field)

Sequence 415 AA;

Query Match 30.7%; Score 629.5; DB 3; Length 415;

Best Local Similarity 37.4%; Pred. No. 5.5e-49;

Matches 148; Conservative 76; Mismatches 137; Indels 35; Gaps 1;





DR WPI; 1996-097631/10.  
DR N-PSDB; AAT58555.  
XX  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis.  
XX  
PS Example 1; Page 113-114; 146pp; French.  
XX  
CC The papA gene of S.pristinaespiralis is involved in the biosynthesis of 4  
CC -dimethylamino-L-phenylalanine (DMPAPA), a precursor for pristinamycin  
CC IA. Upstream of the papA gene, on the complementary strand, is the snbA  
CC gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between  
CC these two genes was sequenced and two open reading frames were  
CC identified. The first (pipA) decodes to an amino acid sequence with  
CC homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The  
CC pipA gene product is likely to catalyse the cyclo- deamination of lysine,  
CC leading to production of pipecolic acid. Mutations in the pipA gene were  
CC shown to affect pipecolic acid synthesis but not the synthesis of 3-  
CC hydroxypicolinic acid. The second open reading frame (snbP) could be  
CC decoded to give a product with homology to hydroxylases of the cytochrome  
CC P450 type. Disruption of the pipA and snbP genes can be used to produce  
CC strains of S.pristinaespiralis which are unable to produce the antibiotic  
CC pristinamycin I but which may be able to produce new, modified forms of  
CC it  
XX  
SQ Sequence 398 AA;  
Query Match 29.6%; Score 606.5; DB 2; Length 398;  
Best Local Similarity 38.0%; Pred. No. 6.8e-47;  
Matches 136; Conservative 66; Mismatches 133; Indels 23; Gaps 7;  
Qy 29 WYEMRKDAPVSFDEENOVWVVELYDDVKVVGDKELFSSCMPQ---QTSSIGNSIIN-M 84  
Db 19 WLRMRHHHPVHEDEYG-AFHVRHADVLTVASDPGVYSSQLSLRLPGSQALSEQILSVI 77  
Qy 85 DPKXHTKIRSVYVVKAFTRPVVKMKQWEPRIQETDELIQKFGHSEFDLVHDFSYPLPVI 144  
Db 78 DPNHRTLRLVVSQAFTRPVADLEPRVTELAGQLDADVG-DTDLVADPAYPLPVI 136  
Qy 145 SELGVPSAHMEQKAWSDLLVSTPKDSEAEKAFLEERDKCEE-----LAA 193  
Db 137 AELIGVPPADRTLFRSMDRL-----CMQVADPADMQFGDDADEYQRLVKEPNRAHHA 191  
Qy 194 FFAGIIIEKRNKPDQIIISILVEAETGKLSGELIPFCTLLLVAGNETTNNLISNAMY 253  
Db 192 YLHDHVTDRRARPANDLISALVAARVEGERLTDEQIVFGALLMAGHVSSTMLIGNTVL 251  
Qy 254 SILTEPGVYELSRHPELMPQAVEALFRFAPAPVLRERIAKRDTEIGGHLIKEGDMVLAP 313  
Db 252 CLKDHPRAAARAADRSLIPALIEVLRLRPITVMAKVTTKDTVLACTTIPAGRMVWVPS 311  
Qy 314 VASANRDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLAPLARLEANTALTSLSAFP 371  
Db 312 LLSANHDEQVTFDPDHLDLARE-GRQIAFGHGIHVCIGAPLARLEGRALPDRFP 368

Search completed: May 28, 2004, 23:17:04  
Job time : 225 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:09:49 ; Search time 400 Seconds  
(without alignments)  
12649.010 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgttaaacgcgcg.....gttcogtgtgaaatgtaa 1191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002s:\*

7: geneseqn2003as:\*

8: geneseqn2003bs:\*

9: geneseqn2003cs:\*

10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1189.4	99.9	1191	3	AAA95666 Bacillus
2	1186.2	99.6	1204	3	AAA95667 Bacillus
3	1181.4	99.2	1221	3	AAA95668 Bacillus
4	1168.6	98.1	1221	3	AAA95669 Bacillus
5	425	35.7	966	6	ABK74899 Bacillus
6	242	20.3	1213	6	ABK74891 Bacillus
7	149	12.5	1233	7	AD55813 Micromono
8	149	12.5	60196	7	AD55810 Micromono
9	139.2	11.7	8478	2	AAQ81792 B. subtil
10	120.4	10.1	1215	8	ACF06128 Bacterial
11	118.2	9.9	110000	4	AAI99682 39
12	116.4	9.8	110000	4	AAI99683 39
13	116.4	9.8	110000	4	AAI99682 39
14	116.4	9.8	110000	4	AAI99683 39
15	114.2	9.6	1248	8	AAH65582 Bacterial
16	112.8	9.5	1290	5	AAH65582 C glutami
17	112.8	9.5	1407	4	AAAF1749 Corynebac
18	112.8	9.5	1414	9	ADD13338 C. glutam
19	112.8	9.5	34980	5	AAH68525 C. glutami
20	111.4	9.4	1194	2	AAT58555 Streptomy
21	111.4	9.4	1257	4	AAF81357 Quorum se
22	111.4	9.4	4496	2	AAT58553 Streptomy
23	110.2	9.3	1155	6	ABK74894 Bacillus

24	106	8.9	1215	8	ACF06133 Bacterial
25	105.4	8.8	8293	6	ABX09140 Mycobacte
26	105.4	8.8	110000	4	AAI99682 21
27	105.4	8.8	110000	4	AAI99683 21
28	103.6	8.7	6085	2	AAT70153 S.longisp
29	103	8.6	215	6	ABK74951 Bacillus
30	102.8	8.6	36538	7	ABV75558 Saccharop
31	98.2	8.2	5880	4	AAI68687 Pseudomon
32	98.2	8.2	11355	4	AAI68692 Pseudomon
33	97.4	8.2	2168	2	AAQ73674 Mycinamic
34	97.2	8.2	110000	4	AAI99682 08
35	97.2	8.2	110000	4	AAI99683 08
36	96.4	8.1	1188	9	ADC36141 Weed cont
37	96.4	8.1	1418	9	ADC36144 Weed cont
38	96.4	8.1	1418	9	ADC36202 Weed cont
39	96.2	8.1	8169	2	AAV26609 Actinomad
40	96	8.1	1191	6	ABN88910 Streptomy
41	95.4	8.0	1278	8	ACF06118 Bacterial
42	93.8	7.9	1236	7	ACF69188 Phototrab
43	93.8	7.9	110000	7	ACF67367 20
44	93.8	7.9	110000	7	ACF67367 21
45	93.8	7.9	243072	7	ACF65382 Phototrab

## ALIGNMENTS

### RESULT 1

AAA95666

ID AAA95666 standard; DNA; 1191 BP.

AC AAA95666;

XX AAA95666;

DT 14-FEB-2001 (first entry)

XX Bacillus subtilis hydroxylating enzyme gene.

DE Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;

XX Hydroxymethylglutaryl co-enzyme A reductase; hypochlosterolaemia.

KW Bacillus subtilis.

XX WO200044886-A1.

PN 03-AUG-2000.

PD 28-JAN-2000; 2000WO-JP000472.

XX 29-JAN-1999; 99JP-C0021707.

PR (KYOW ) KYOWA HAKKO KOGYO KK.

XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;

XX WPI; 2000-548827/50.

DR P-PSDB; AAB15501.

XX New protein derived from Bacillus genus microorganism useful for

PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase

XX inhibitors.

PS Claim 14; Page 64-67; 111pp; Japanese.

CC This sequence represents the coding region for a novel Bacillus derived protein having the activity of producing a hydroxylated bicyclic compound or the corresponding lactone from a bicyclic compound or the corresponding lactone. The protein is used for preparing hydroxylated compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and hypocholesterolaemic agents

XX Sequence 1191 BP; 330 A; 248 C; 311 G; 302 T; 0 U; 0 Other;

SQ Query Match 99.9%; Score 1189.4; DB 3; Length 1191;

Best Local Similarity 99.9%; Pred. No. 0; Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY 1 ATGAATGTGTTAAACCGCGCAAGCCTTGCAGCGAGCGTCTCAATGGGAAAAACAAA 60	1021 GCCTTTGGCCACGCAATCCATTTTTCCTTGGGCGCCCGCTTGCCTTGAAGCAAAAT 1080	
DB 1 ATGAATGTGTTAAACCGCGCGCAAGCCTTGCAGCGAGCGTCTCAATGGGAAAAACAAA 60	QY 1081 ATCGCGTTAAACGCTTTTGATTTCTGCTTTTCTCATATGAGTGGTCACTATCCTCCG 1140	
QY 61 CAGGATGCGTATCATCGCTTTCATCGTATGAATCATGAGAAAGATGCGCTGTTTCC 120	DB 1081 ATCGCGTTAAACGCTTTTGATTTCTGCTTTTCTCATATGAGTGGTCACTATCCTCCG 1140	
DB 61 CAGGATGCGTATCATCGCTTTCATCGTATGAATCATGAGAAAGATGCGCTGTTTCC 120	QY 1141 ATTGAAAAACAGTGTGATATACCGATTAAAGAGCTTCGGTGTGAAAAATGTAA 1191	
QY 121 TTTCATGAAGAAAAACCAAGTGTGAGCGCTTTCTTTATGATGATCTCAAAAAGTTGTT 180	DB 1141 ATTGAAAAACAGTGTGATATACCGATTAAAGAGCTTCGGTGTGAAAAATGTAA 1191	
DB 121 TTTCATGAAGAAAAACCAAGTGTGAGCGCTTTCTTTATGATGATCTCAAAAAGTTGTT 180	RESULT 2	
QY 181 GGGGATAAAGAGTGTGTTTCCAGTTGCATGCCGACAGACAGCAAGCTCTATTGGAATTC 240	AAA95667	
DB 181 GGGGATAAAGAGTGTGTTTCCAGTTGCATGCCGACAGACAGCAAGCTCTATTGGAATTC 240	ID AAA95667 standard; DNA; 1204 BP.	
QY 241 ATCATTAACNTGGACCGCGGAGCAGCATACAAAATCGTTTCAGTCTGTGAACAAAGCCTTT 300	XX AAA95667;	
DB 241 ATCATTAACNTGGACCGCGGAGCAGCATACAAAATCGTTTCAGTCTGTGAACAAAGCCTTT 300	AC AC	
QY 301 ACTCGCGCGTGCATGAAGCAATGGGAACCGAGAATTCAGAAATCACAGATGAACCTGATT 360	XX 14-FEB-2001 (first entry)	
DB 301 ACTCGCGCGTGCATGAAGCAATGGGAACCGAGAATTCAGAAATCACAGATGAACCTGATT 360	XX Bacillus subtilis hydroxylating enzyme gene #2.	
QY 361 CAAAATTTTCAGGGCGAGTGGTTGACCTTTGTCAGATTTTTCATACCGGCTTCGG 420	DE Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;	
DB 361 CAAAATTTTCAGGGCGAGTGGTTGACCTTTGTCAGATTTTTCATACCGGCTTCGG 420	XX Hydroxymethylglutaryl co-enzyme A reductase; hypcholesterolaemia.	
QY 421 GTTATGTGATATCTGAGCTGTGGAGTGGCTTCAGCGACATGGAAAGCTTTAAAGCA 480	XX OS	
DB 421 GTTATGTGATATCTGAGCTGTGGAGTGGCTTCAGCGACATGGAAAGCTTTAAAGCA 480	XX Bacillus subtilis.	
QY 481 TGGTCTGATCTTCTGGTCAGTACACCGAAGGATAAAAGTGAAGAACTGAAAAGCCTTT 540	XX PN	WO200444886-A1.
DB 481 TGGTCTGATCTTCTGGTCAGTACACCGAAGGATAAAAGTGAAGAACTGAAAAGCCTTT 540	XX PD	03-AUG-2000.
QY 541 TTGGAGAACGAGATAGTGTGAGGAGAACTGGCGCGCTTTTTCGCGGATCATAGAA 600	XX PF	28-JAN-2000; 2000WO-JP000472.
DB 541 TTGGAGAACGAGATAGTGTGAGGAGAACTGGCGCGCTTTTTCGCGGATCATAGAA 600	XX PR	29-JAN-1999; 99JP-00021707.
QY 601 GAAAGCGAAAACAAACCGGAAACAGGATATTTCTATTTAGTGGAGCGGAAAGCA 660	XX PA	(KYOW ) KYOWA HAKKO KOGYO KK.
DB 601 GAAAGCGAAAACAAACCGGAAACAGGATATTTCTATTTAGTGGAGCGGAAAGCA 660	XX PI	Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
QY 661 GCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTTCAGCGCTGCTGCTGCGCGGA 720	XX PP	WPI; 2000-548827/50.
DB 661 GCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTTCAGCGCTGCTGCTGCGCGGA 720	DR P-PSDB; AAB15502.	
QY 721 AATGAACCACTACAAACCTGATTTCAATGGCATGTACAGCATATTAAGAACCGCAGGC 780	XX PT	New protein derived from Bacillus genus microorganism useful for
DB 721 AATGAACCACTACAAACCTGATTTCAATGGCATGTACAGCATATTAAGAACCGCAGGC 780	XX PT	producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
QY 781 GTTTACAGAGAACTGCGCAGCCTCTCACTGATGCTTCAGCGAGTGGAGAACTTG 840	XX PS	inhibitors.
DB 781 GTTTACAGAGAACTGCGCAGCCTCTCACTGATGCTTCAGCGAGTGGAGAACTTG 840	XX Claim 16; Page 90-94; 111pp; Japanese.	
QY 841 CGTTTCAGAGCGCGCGCGCTTTTGGAGCGCATTTGCCAAGCGGATAACGAGATCGGG 900	XX This sequence represents the coding region for a novel Bacillus derived	
DB 841 CGTTTCAGAGCGCGCGCGCTTTTGGAGCGCATTTGCCAAGCGGATAACGAGATCGGG 900	CC protein having the activity of producing a hydroxylated bicyclic compound	
QY 901 GGGCACTGATTAAGAGAGTGTATGTTTGGCGTTTGGCGCTTTGGCGCTTTGGCGCTTT 960	CC or the corresponding lactone from a bicyclic compound or the	
DB 901 GGGCACTGATTAAGAGAGTGTATGTTTGGCGTTTGGCGCTTTGGCGCTTTGGCGCTTT 960	CC corresponding lactone. The protein is used for preparing hydroxylated	
QY 961 GAAGCAAGTTTTCAGACCGCAGATGTTTGATATCCCGCGCATCCCAATCCGATATT 1020	CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase	
DB 961 GAAGCAAGTTTTCAGACCGCAGATGTTTGATATCCCGCGCATCCCAATCCGATATT 1020	XX inhibitors and hypcholesterolaemic agents	
QY 1021 GCGTTTGCCCAACGCAATCCATTTTTCCTTGGGCGCGCTTGGCGCTTGAAGCAAAAT 1080	QY 1 ATGATGTGTTAAACCGCGCAAGCCTTGCAGCGAGCGTCTCAATGGGAAAAACAAA 60	
DB 1021 GCGTTTGCCCAACGCAATCCATTTTTCCTTGGGCGCGCTTGGCGCTTGAAGCAAAAT 1080	DB 8 ATGATGTGTTAAACCGCGCAAGCCTTGCAGCGAGCGTCTCAATGGGAAAAACAAA 67	
	QY 61 CAGGATGCGTATCATCGCTTTCATCGTATGAATCATGAGAAAGATGCGCTGTTTCC 120	
	DB 68 CAGGATGCGTATCATCGCTTTCATCGTATGAATCATGAGAAAGATGCGCTGTTTCC 127	
	QY 121 TTTCATGAAGAAAAACCAAGTGTGAGCGCTTTCTTTATGATGATGTCAAAAAGTTGTT 180	
	DB 128 TTTCATGAAGAAAAACCAAGTGTGAGCGTTTTCTTTATGATGATGTCAAAAAGTTGTT 187	
	QY 191 GGGGATAAAGAGTGTGTTTTCAGTTGCATGCCGACAGACAGCTCTATTGGAATTC 240	

```
Db 188 GGGGATAAAGAGTGTGTTTCCAGTTGCGCGAGCAGACAGCTCTATTGGAAATCC 247
Qy 241 ATCAATTAACATGACCGCCGCGAGCATACAAATCCCGTTCACTGCTGAAACAAAGCCTTT 300
Db 248 ATCAATTAACATGACCGCCGCGAGCATACAAATCCCGTTCACTGCTGAAACAAAGCCTTT 307
Qy 301 ACTCCGCGGTGATGAAGCAATGGGAACCGGAGATTCAGAAATCAAGATCAAGATCAAGT 360
Db 308 ACTCCGCGGTGATGAAGCAATGGGAACCGGAGATTCAGAAATCAAGATCAAGATCAAGT 367
Qy 361 CAAAAATTTACGGGCGCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 368 CAAAAATTTACGGGCGCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
Qy 421 GTTATTGTGATATCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTG 480
Db 428 GTTATTGTGATATCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTG 487
Qy 481 TGGTCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 488 TGGTCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
Qy 541 TTGGAAGAAACGAGATAAGTGTGAGAGAACTGGCGCGCTTTTTCGCGGCTCATAGAA 600
Db 548 TTGGAAGAAACGAGATAAGTGTGAGAGAACTGGCGCGCTTTTTCGCGGCTCATAGAA 607
Qy 601 GAAAGCGAAACAAACCGGAAACAGGATATTTTCTATTTTGTAGTGAACGCGAAGAAACA 660
Db 608 GAAAGCGAAACAAACCGGAAACAGGATATTTTCTATTTTGTAGTGAACGCGAAGAAACA 667
Qy 661 GGGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 668 GGGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
Qy 721 AATGAACACCACTACAAACCTGATTTCAATCGCATGTACAGCATATTAGAAACCCAGGC 780
Db 728 AATGAACACCACTACAAACCTGATTTCAATCGCATGTACAGCATATTAGAAACCCAGGC 787
Qy 781 GTTTACGAGAACTGCGCAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 788 GTTTACGAGAACTGCGCAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 847
Qy 841 CGTTTCAGAGCGCGCGCGCGCTTTGAGGCGCATTCGCAAGCGGGATACGGAGATCGGG 900
Db 848 CGTTTCAGAGCGCGCGCGCGCTTTGAGGCGCATTCGCAAGCGGGATACGGAGATCGGG 907
Qy 901 GGGCAGCTGATTAAGAGAGTGATATGTTTGGCGTTTGGCGATCGGCAAAATCGTGAT 960
Db 908 GGGCAGCTGATTAAGAGAGTGATATGTTTGGCGTTTGGCGATCGGCAAAATCGTGAT 967
Qy 961 GAAGCAAAAGTTTGACAGACGCGCACTGTTTGAATATCCGCGCATCCCAATCCGCATATT 1020
Db 968 GAAGCAAAAGTTTGACAGACGCGCACTGTTTGAATATCCGCGCATCCCAATCCGCATATT 1027
Qy 1021 GCGTTTGGCAGCGCATCCATTTTGGCTTGGGCGCGCGCTTGGCGCGCTTGGAGCAAAAT 1080
Db 1028 GCGTTTGGCAGCGCATCCATTTTGGCTTGGGCGCGCGCTTGGCGCGCTTGGAGCAAAAT 1087
Qy 1081 ATCGCGTTTAACTGCTTTGATTTCTGCTTTTCTCATATGAGTGGCGTCACTACCTCG 1140
Db 1088 ATCGCGTTTAACTGCTTTGATTTCTGCTTTTCTCATATGAGTGGCGTCACTACCTCG 1147
Qy 1141 ATTGAAGAAACAGTGTGATATACGAGTTAAGAGCTTCCGTTGTGAAATGTAA 1191
Db 1148 ATTGAAGAAACAGTGTGATATACGAGTTAAGAGCTTCCGTTGTGAAATGTAA 1198
```

RESULT 3  
AAA95668  
ID AAA95668 standard; DNA; 1221 BP.  
XX  
AC AAA95668;

```
XX 14-FEB-2001 (first entry)  
XX Bacillus subtilis hydroxylating enzyme gene #3.  
XX DE Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;  
XX KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.  
XX OS Bacillus subtilis.  
XX PN W0200044886-A1.  
XX PD 03-AUG-2000.  
XX PF 28-JAN-2000; 2000WO-JP000472.  
XX PR 29-JAN-1999; 39JP-00021707.  
XX PA (KYOW) KYOWA HAKKO KOGYO KK.  
XX PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX WPI; 2000-548827/50.  
XX New protein derived from Bacillus genus microorganism useful for  
XX producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
XX inhibitors.  
XX Claim 16; Page 97-100; 111pp; Japanese.  
XX This sequence represents the nucleotide sequence for a novel Bacillus  
XX derived protein having the activity of producing a hydroxylated bicyclic  
XX compound or the corresponding lactone from a bicyclic compound or the  
XX corresponding lactone. The protein is used for preparing hydroxylated  
XX compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase  
XX inhibitors and hypocholesterolaemic agents  
XX SQ Sequence 1221 BP; 336 A; 257 C; 320 G; 308 T; 0 U; 0 Other;
```

```
Query Match 99.2%; Score 1181.4; DB 3; Length 1221;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1185; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 ATGAATGTGTAAACCGCGCGCAAGCCTTGCGAGCGAGCGCTGCTCAATGGGAAACAA 60  
Db 25 ATGAACGTCTGAACCGCGCTCAAGCCTTGCGAGCGCGCTGCTCAATGGGAAACAA 84  
Qy 61 CAGGATCGGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGGATCGCCTGTTTC 120  
Db 85 CAGGATCGGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGGATCGCCTGTTTC 144  
Qy 121 TTTCATGAAGAAACCAAGTGTGAGCGTCTTCTTCTTATGATGATGTCAGAAAGTGT 180  
Db 145 TTTCATGAAGAAACCAAGTGTGAGCGTCTTCTTCTTATGATGATGTCAGAAAGTGT 204  
Qy 181 GGGGATTAAGAGTGTGTTTCCAGTTGCTGCGCGAGCAGCAAGCTCTATTGGAAATCC 240  
Db 205 GGGGATTAAGAGTGTGTTTCCAGTTGCTGCGCGAGCAGCAAGCTCTATTGGAAATCC 264  
Qy 241 ATCAATTAACATGACCGCGCGAGCATACAAATCCCGTTCACTGCTGAAACAAAGCCTTT 300  
Db 265 ATCAATTAACATGACCGCGCGAGCATACAAATCCCGTTCACTGCTGAAACAAAGCCTTT 324  
Qy 301 ACTCCGCGGTGATGAAGCAATGGGAACCGGAGATTCAGAAATCAAGATCAAGATCAAGT 360  
Db 325 ACTCCGCGGTGATGAAGCAATGGGAACCGGAGATTCAGAAATCAAGATCAAGATCAAGT 384  
Qy 361 CAAAAATTTACGGGCGCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
Db 385 CAAAAATTTACGGGCGCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444  
Qy 421 GTTATTGTGATATCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTGCTGAGTCTG 480
```



Db	445	GTATTGTCGATATCTGAGCTGCTGGGAGTGCCCTTCACGCGATATGGAAACAGTTTAAAGCA	504
Qy	481	TGGTCTCATCTCTCTGCTCAGTACACCGAAGATAAAAGTGAAGAGCTGAAAAGCCCTTT	540
Db	505	TGGTCTCATCTCTCTGCTCAGTACACCGAAGATAAAAGTGAAGAGCTGAAAAGCCCTTT	564
Qy	541	TTGGAAGAAACGAGATAAGTGTGAGGAAGAACTGSCCGCGTTTTTTTGCCGCATCATAGAA	600
Db	565	TTGGAAGAACGAGATAAGTGTGAGGAAGAACTGSCCGCGTTTTTTTGCCGCATCATAGAA	624
Qy	601	CAAAAGCGAAACAAACCGGAACGAGATATTATTCTATTATTTAGTGGAGCGGAGAAACA	660
Db	625	CAAAAGCGAAACAAACCGGAACGAGATATTATTCTATTATTTAGTGGAGCGGAGAAACA	684
Qy	661	GGCAGAGAGCTGCCGGTGAAGAGCTCATTCGGTTTTTTTTCACGCTGCTGCTGGGCCGGA	720
Db	685	GGCAGAGAGCTGCCGGTGAAGAGCTCATTCGGTTTTTTTTCACGCTGCTGCTGGGCCGGA	744
Qy	721	AATGAAACCACTACAAACCTGATTTCAATATGCGATGTACAGCATATTAGAAACGCCAGGC	780
Db	745	AATGAAACCACTACAAACCTGATTTCAATATGCGATGTACAGCATATTAGAAACGCCAGGC	804
Qy	781	GTTTACAGGAACTGCGCAGCATCTCTGAACTGATGCTCAGGCAATCGAGATCGGG	840
Db	805	GTTTACAGGAACTGCGCAGCATCTCTGAACTGATGCTCAGGCAATCGAGATCGGG	864
Qy	841	CGTTTCAGAGCGCGCGCCCGTTTTTCAGCGCGCATTCGCAAGCGGGATACGGAGATCGGG	900
Db	865	CGTTTCAGAGCGCGCGCCCGTTTTTCAGCGCGCATTCGCAAGCGGGATACGGAGATCGGG	924
Qy	901	GGGCACCTGATTTAAAGAAAGTGATATGTTTTTGCGGTTTGTGGCATCGGCAAAATCGTGAT	960
Db	925	GGGCACCTGATTTAAAGAAAGTGATATGTTTTTGCGGTTTGTGGCATCGGCAAAATCGTGAT	984
Qy	961	GAAGCAAAAGTTTGACAGACCGCATGTTTGATATCGCGCGCATCCCAATCCGCATATT	1020
Db	985	GAAGCAAAAGTTTGACAGACCGCATGTTTGATATCGCGCGCATCCCAATCCGCATATT	1044
Qy	1021	GGGTTTGGCCACGGCATCCATTTTTCCTTTGGGGCCCGCTTGGCCGCTCTTGGAGCAAT	1080
Db	1045	GGGTTTGGCCACGGCATCCATTTTTCCTTTGGGGCCCGCTTGGCCGCTCTTGGAGCAAT	1104
Qy	1081	ATCGCGTTAAAGCTCTTGATTTCTGTTTTTCTCATATGGAGTGGCGTCAGTATCACTCCG	1140
Db	1105	ATCGCGTTAAAGCTCTTGATTTCTGTTTTTCTCATATGGAGTGGCGTCAGTATCACTCCG	1164
Qy	1141	ATTGAAACCAAGTGTGATATACGGATTTAAAGAGCTTCCTGCTGGAATAATGTAA	1191
Db	1165	ATTGAAACCAAGTGTGATATACGGATTTAAAGAGCTTCCTGCTGGAATAATGTAA	1215

## RESULTS

RESULT 4  
AAA95669

AAA95669  
ID AAA95669 standard: DNA: 1221 BP.

ID  
 YY  
 AAH95069

XX  
AC  
AAA95669:

AC  
XX  
XX  
AAH93009;

DT 14-FEB-2001 (first entry)

DI 14-FEB-11 11:18 AM (TUE)

DE *Bacillus subtilis* hydroxylating enzyme gene #4.

XX  
XX

KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;

KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.

XX

OS Bacillus subtilis.

XX

XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX PA  
XX PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX  
XX DR WPI; 2000-548827/50.  
XX DR P-PSDB; AAB15503.  
XX  
XX PT New protein derived from Bacillus genus microorganism useful for  
XX PT producing a hydroxylated bicyclic compound e.g. HM-CoA reductase  
XX PT inhibitors.  
XX  
XX PS Claim 16; Page 101-104; 11pp; Japanese.  
XX  
XX CC This sequence represents the coding region for a novel Bacillus derived  
XX CC protein having the activity of producing a hydroxylated bicyclic compound  
XX CC or the corresponding lactone from a bicyclic compound or the  
XX CC corresponding lactone. The protein is used for preparing hydroxylated  
XX CC compounds useful as hydroxymethylglutaryl co-enzyme A (HM-CoA) reductase  
XX CC inhibitors and hypocholesterolaemic agents  
XX  
XX SQ Sequence 1221 BP; 334 A; 261 C; 322 G; 304 T; 0 U; 0 Other;

...

Db 685 GCGAGAGAGCTGTCGGGTGAAGAGCTGATTCGGTTGTGACGCTGCTGCTGTCGGTGGCCGGA 744  
Qy 721 AATGAACACCACTCAAAACCTGATTTCAATTCGCGATGTACAGCATATTAGAAACCCAGGC 780  
Db 745 AATGAACACCACTCAAAACCTGATTTCAATTCGCGATGTACAGCATATTAGAAACCCAGGC 804  
Qy 781 GTTTAGAGAACTGCGCAGCATCTCTGACTGATGCTTACAGGAGTGGAGAGCCTTG 840  
Db 805 GTTTAGAGAACTGCGCAGCATCTCTGACTGATGCTTACAGGAGTGGAGAGCCTTG 864  
Qy 841 CGTTTCAGAGCGCGCGCCCGCTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGG 900  
Db 865 CGTTTCAGAGCGCGCGCCCGCTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGG 924  
Qy 901 GGGCACCTGATTAAGAAGGTGATGCTTTTGGCGTTTGTGGCATCGGCAATTCGTGAT 960  
Db 925 GGGCACCTGATTAAGAAGGTGATGCTTTTGGCGTTTGTGGCATCGGCAATTCGTGAT 984  
Qy 961 GAAGCAAGTTTGACACCGCACATGTTTGATATCGCGCCCATCCCATCCGCATATT 1020  
Db 985 GAAGCAAGTTTGACACCGCACATGTTTGATATCGCGCCCATCCCATCCGCATATT 1044  
Qy 1021 GCGTTTCGCGCAGCATCCATTTTTCCTTGGGCGCCCGCTTGGCGCTTTGAAGCAAT 1080  
Db 1045 GCGTTTCGCGCAGCATCCATTTTTCCTTGGGCGCCCGCTTGGCGCTTTGAAGCAAT 1104  
Qy 1081 ATCGGTTAAACGTTTGAATTCCTGCTTTTCTCATATGAGAGTGGTCACTCCG 1140  
Db 1105 ATCGGTTAAACGTTTGAATTCCTGCTTTTCTCATATGAGAGTGGTCACTCCG 1164  
Qy 1141 ATTGAACAGTGTGATATACGGATTAAGAGCTTCGGTGAAGTAA 1191  
Db 1165 ATTGAACAGTGTGATATACGGATTAAGAGCTTCGGTGAAGTAA 1215

## RESULT 5

ABK74899

ID ABK74899 standard; DNA; 966 BP.

AC ABK74899;

XX

XX

DT 13-AUG-2002 (first entry)

DE Bacillus licheniformis genomic sequence tag (GST) #2190.

XX

KW Differential gene expression; genomic sequenced tag; GST;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

XX

OS Bacillus licheniformis.

XX

PN W0200229113-A2.

XX

PD 11-APR-2002.

XX

PF 05-OCT-2001; 2001MO-US031437.

XX

PR 06-OCT-2000; 2000US-00680598.

XX

PR 27-MAR-2001; 2001US-0279526P.

XX

XX (NOVO ) NOVOZYMES BIOTECH INC.

XX

PA (NOVO ) NOVOZYMES AS.

XX

PI Berka R, Clausen IG;

XX

XX WPI; 2002-416684/44.

XX

PT Monitoring differential expression of several genes in first Bacillus

PT cell relative to expression of same genes in one or more second Bacillus

PT cells, by using substrate containing Bacillus genomic sequenced tag

PT array.

XX

XX Claim 4; SEQ ID NO 2190; 200pp; English.

PS

XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 966 BP; 252 A; 229 C; 269 G; 216 T; 0 U; 0 Other;

Query Match 35.7%; Score 425; DB 6; Length 966;

Best Local Similarity 65.4%; Pred. No. 7.1e-127;

Matches 623; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 230 TTGGAATTCCTCATTAACATGACCGCGCGAGCATACAAATCCGTTCAAGTCGTA 289

Db 7 TTGGAAACGATCATGATGATCCCGGAAACACACGAGATCAGATCGTCA 66

Qy 290 ACAAAGCCTTTACTCCGCGCTGATGAAGCAATGGAACCGAGATTCAGAAATCACAG 349

Db 67 ACAAGGCTTTTACACCGCGCTGATGAAGAGTGGAGCGCGCATTCGGAATCGA 126

Qy 350 ATGAACGATTCAAATAATTTACGGGGCGCATGAGTTGACCTTGTTCACGATTTTCA 409

Db 127 ATCAATTCCTGCTGATGTCGCGCGGAGGAGATCGACCTTGTACAGGATTTTCA 186

Qy 410 ACCGCTTCGGTTATTGTGATATCTGAGTCTGGGAGTGCCTTCAGCGAGATGGAC 469

Db 187 ATCCGTTGCCGTTAATGCTCAATTCGGAATTCCTCGGCTTCCTTGTGTATAGATC 246

Qy 470 AGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAGGATAAAAGTGAAGCTG 529

Db 247 ATTTCAAGAAATGCTCCGACCTGTTGTCAGCTGCCGAGAGCGATCGCGCGAAGATG 306

Qy 530 AAAAAAGCCTTTTGGAGAACGATATAGTGTGAGGAGAACTGCGCGGTTTTTGGCG 589

Db 307 TGAAACGAGTGGAAAAACATCAGGGACCAAGCGAGAGAGCTGACCGCATTCITGAAA 366

Qy 590 GCATCATAGAGAAAGCGAAACAAACCGAAACAGGATATTATTCTATTATTAGTGAAG 649

Db 367 AGATGTTGAGAGAAAGACAAACCTCGCATGACCTGATTTGCTTTGATCAAG 426

Qy 650 CGGAAGAAACAGGCGAGAGCTGTCCGGTGAAGAGCTGATTCGTTTTGACGCTGTC 709

Db 427 CCGBACAGAGAGGAGCAGCTGTCCCTGATGAATGGTTCCATTTTGAATCTGCT 486

Qy 710 TGGTGGCGGAAATGAACCACTACAACTGATTTTGAATGCGATGATGAGCATATTAG 769

Db 487 TGA TGGCGGGGAAATGAACCACTGATTTAGTTTCAATTCGCTCTACAGCATTCCTCG 546

Qy 770 AAACGCGAGCGGTTTACGAGAACTGCGGCGCATCTGACTGATGCTCAGGAGTGG 829

Db 547 AAACCGCGGCGTGTATGACGAGCTCGCCAGGATCTTGAATCTCGCAGGAGTGG 606

Qy 830 AGGAAGCTTTCGTTTTCAGAGCGCGCGCGGTTTTGAGCGCATTCGCCAAGCGGATA 889

Db 607 AGGAAGCGCTCCGTTTCCGGCGCGCAGCGCGATGATGTGCTTGTAAACAGGATA 666



Db 926 AAGGTTCTACAGTATGCTGGGTGGTTCAGCGAATCGTGACGAATTAAGTTGA 985  
 QY 975 CAGACCGCACATGTTGATATCCGCGCCATCCCAATCCGCATATTGCGTTGGCCACGG 1034  
 Db 986 CGATCCTGACAGCTTCAAGCTTGATCGCAATCAAACTTCATAGAGCTTCGGCTGG 1045  
 QY 1035 CATCCATTTTGGCTTGGGCGCCGCTTGGCCGCTTGAAGCAAAATATCGGTTAAGTC 1094  
 Db 1046 CATCCATTTTGGCTTGGGCGCCGCTTGGCCGCTTGAAGCAAAATATCGGTTAAGTC 1105  
 QY 1095 TTGAT 1100  
 Db 1106 CTGCT 1111

## RESULT 7

AAD5813  
 ID AAD5813 standard; DNA; 1233 BP.

XX AAD5813;

XX 27-OCT-2003 (revised)

DT 07-AUG-2003 (first entry)

XX Micromonospora carbonacea polyketide synthase (PKS) type I gene #3.

KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme; gene; ds.

XX Micromonospora carbonacea.

XX Key Location/Qualifiers

EH 1.1233

FT CDS

FT /tag= a

FT /product= "Polyketide synthase"

XX CA2391131-A1.

XX 19-NOV-2002.

XX 26-JUL-2002; 2002CA-02391131.

XX 26-JUL-2001; 2001US-0307629P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Yang X, Staffa A, Farnet CW;

XX WPI; 2003-343556/33.

DR P-PSDB; AAE36997.

XX Novel isolated polypeptide involved in biosynthesis of macrolides by microorganisms, useful for biosynthesis of macrolides by microorganisms, preferably for biosynthesis of rosaramicin.

XX Claim 1; Page 101-102; 206pp; English.

XX The invention relates to genes and proteins involved in the biosynthesis of macrolides by microorganisms. In particular it relates to the nucleic acids forming the biosynthetic locus for rosaramicin (a 16-member macrolide antibiotic) from Micromonospora carbonacea. The invention is useful for the biosynthesis of macrolides by microorganisms. It allows direct manipulation of macrolides and related chemical structures by chemical engineering of the proteins involved in the biosynthesis of rosaramicin. It is useful to catalyse certain biochemical reactions, in vitro or in vivo, to direct or enhance the synthesis or modification of a polyketide, polyketide substrate or its precursor. The present sequence is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-OCT-2003 to standardise OS field)

XX Sequence 1233 BP; 225 A; 452 C; 377 G; 179 T; 0 U; 0 Other;

XX Query Match 12.5%; Score 149; DB 7; Length 1233;

Best Local Similarity 48.5%; Pred. No. 4.2e-37;  
 Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;  
 QY 239 CCATCAATTAACATGAGACCCCGCGAAGCATACAAAATCCGTTGAGTGGTGAACAAAGCCT 298  
 Db 254 CCATCGCGGCATCGACCCCGCGCGGCGCGCGCTGCGCAAGCTGTGAGCCAGCGGT 313  
 QY 299 TTACTCCGCGGTGATGAAGCAATGGGAACCGAGAATTCAGAAATCACAGATGAATGA 358  
 Db 314 TCACCCCGCGCGATCGCCAGATGAGCTGCGGATCGGGCAGATCACCGCCAGCGTC 373  
 QY 353 TTCAAAATTTACGGGCGCAGTTCAGTTTCACCTGTTTCAGATTTTTCATACCCGCTTC 418  
 Db 374 TCAGACAGGTACCGACCGAGACCGGATCGACATCGCCAGCGCTCGGTACCGCTGC 433  
 QY 419 CGGTATTGTATATCTGAGCTCTGGAGTGGCTTCAGCGCGAGATGGAACAGTTTAAAG 478  
 Db 434 CGGTGAGGTCTATCGCGAGCTGCTCGGCATTCACCAAGGATCAGAGNAGTTCCGG 493  
 QY 479 CATGCTCTGATCTTCTGGTCACTACCGAAGGATAAAGTGAAGAGCTGAAAAGCCT 538  
 Db 494 AGTGGGTGGACATCATCTCAGCAACGAAGGCTGGAGTATCCCAAGCTCCCGACGACT 553  
 QY 539 TTTTGGAGACCGAGATAAGTGTGAGGAAGTCTGGCGGCTTTTTCGGGATCATAG 598  
 Db 554 TCACCGAGAGCGTGGCGCCCGCCATCGAGAGTGTCCGAATTCCTGTACGCCAGATCG 613  
 QY 599 AAGAAAAGCGAAACAAACCGAACAGGATATATTCTATTATTAGTGAACCGGAAGAA 658  
 Db 614 CCCACAGCGCGGACCGAACCGAGACCTGTATCAGCGCTCTGTGCGCGAGGTGCG 673  
 QY 659 CAGCGAGAGAGTGTCCGGTGAAGAGCTGATTCGGTTTTCACCGCTGTGTGTGGCGG 718  
 Db 674 ACGGCGCAAGCTGACCGAGAGAGTGTCTCAACATCGTCGCGCTGTCTGCTCAACGCG 733  
 QY 719 GAATGAACCACTACAAACCTGATTTCAATGGATGTACAGCATATTAGAACCGCAG 778  
 Db 734 GGCACATCTCCAGCGCACGCTGCTCAGCAACCTGTTCTGCTGTGAGGAGCACCGCG 793  
 QY 779 GCGTTTACAGGAAGTCTCGCAGCGCATCTGAACTGATGCTTCAAGGAGTGAAGAGCCT 838  
 Db 794 AGGCACAGCGCGGTTCGCGCGCGACCGCAGCTCTGTGCGGGGTGNTCCAGAGGAGCG 853  
 QY 839 TGGTTTTCAGAGCGCGCGCGCGGTTCAGGCGCATTCGCAAGCGGAGTACGAGATCG 898  
 Db 854 TGGCTACCGGTCCCGCTTCAACTGCACTCTTCGGATCTCTGAACGAGGACACCGACATCC 913  
 QY 899 GGGGCGACCTGATTAAGAGAGTGTATGTTTGGGCTTTGTGGCATTCGGCAATCGTG 958  
 Db 914 TCGGCCACCCCATCGCGAAGGCGAGATGATGATCGCTCGCTGATCGCTTCGCGAACCGCG 973  
 QY 959 ATGAAGCAAAAGTTTGACAGACCGCACATGTTGATATCCGCGCGCATCCCAATCCGATA 1018  
 Db 974 ACACCGAGGTGTTACCGACCGCGACACTTCGACATCCGACCGAGTTCGACCAAGCACC 1033  
 QY 1019 TTGCGTTTGGCCAGCGCATTCATTTTTCCTTGGGCGCCCGCTTGGCGCTTCTTAAGCAA 1078  
 Db 1034 TGGCGTTTGGCCAGCGCATTCACCTTGGGCGCGGTTCCTGGCCAGGCTGGAGGCGA 1093  
 QY 1079 ATATC 1083  
 Db 1094 AGGTC 1098

## RESULT 8

AAD5810/c

ID AAD5810 standard; DNA; 60196 BP.

XX AAD5810;

XX 27-OCT-2003 (revised)

DT 07-AUG-2003 (first entry)

XX



```

Db 3248 CCACAAAGCGCGCGAACCAGACACACCTGATCAGCGCCTCTGTGCGCGGAGGTGG 3189
QY 659 CAGCGCAGAGCTGTCGGTGAAGAGCTGATTCGGTTTTCACGCTGCTGCTGTGCGCG 718
Db 3188 ACGGCGCAGCTGACCGAGCAGAGAGTCTCAACATCGTTCGCGCTGCTCTACCGCGC 3129
QY 719 GAAATGAAACCACTACAAACCTGATTTCAAATCGATGTACAGCATATTAGAAACGCCAG 778
Db 3128 GGCACATCTCCAGCGCCAGCTGCTCAGCAACCTGTTCTCTGCTGCTGAGGAGCACCGC 3069
QY 779 GCGTTTACGAGGAACCTCGCGAGCACTCTCAACTGATGCTCTAGGCACTGAGGAGCCT 838
Db 3068 AGGCACAGCGCGCGGTCCGCGCGGACCGCAGCCTCGTGCCTGGCGGTGATCGAGGAGC 3009
QY 839 TCGGTTTCAGAGCGCGCGCGCGGTTTTCAGGCGCATTCGCAAGCGGATACGAGATCG 898
Db 3008 TCGGCTACCGGTCCCGGTTTCACTGCMCTTCGCGATCTCGAGGAGCACCGCATCC 2949
QY 899 GCGGCGACCTGATTAAAGAAAGGTGATATGGTTTGGCGTTTGTGGCATCGGCAAAATCGT 958
Db 2948 TCGGCCACCCATCGCGCAAGGCGAGATGTGATCGCTGATCGCTCCGCAACCGCG 2889
QY 959 ATGAAGCAAGTTTGACAGACCGCACATGTTTGAATCCGCGCCATCCCAATCCGCATA 1018
Db 2888 ACACCGAGGTGTTCAAGGACCGCGACACCTTCGACATCCGACCGAGTCAACCAAGCAC 2829
QY 1019 TTGCGTTTGGCCACCGCATCATTTTTCCTTGGCGCGCGCGCTGCGCGCTTTGAAGCAA 1078
Db 2828 TGGCGTTGGCCACCGCATCCACCACTGCTTGGCGCGGTCTTGGCCAGCTGGAGCGGA 2769
QY 1079 ATATC 1083
Db 2768 AGGTC 2764

```

## RESULT 9

AAQ81792

ID AAQ81792 standard; DNA; 8478 BP.

XX AC AAQ81792;

XX DT 25-MAR-2003 (revised)

XX DT 26-SEP-1995 (first entry)

XX DB B. subtilis biotin operon and flanking sequences.

XX KW Biotin operon; recombinant production; dietary additive; animal feed;

XX KW vitamin supplement; research reagent; ds.

XX OS Bacillus subtilis.

XX FH Key Location/Qualifiers

XX FT terminator /tag= k

XX FT promoter /note= "rho-independent termination site"

XX FT misc\_feature /tag= a

XX FT /note= "sigma A promoter"

XX FT /tag= b

XX FT /note= "potential regulatory site"

XX FT /tag= c

XX FT /product= "bioW"

XX FT /tag= d

XX FT /product= "bioA"

XX FT /tag= e

XX FT /product= "bioF"

XX FT /tag= f

XX FT /tag= i

```

FT CDS /product= "bioD"
FT FT 4408..5415
FT FT /*tag= g
FT FT /product= "bioB"
FT FT 5423..5462
FT FT /*tag= i
FT FT /note= "rho-independent termination site"
FT FT 5484..6671
FT FT /*tag= h
FT FT /product= "bioI"
FT FT 6748..7509
FT FT /*tag= i
FT FT /label= ORF 2
FT FT 7501..7543
FT FT /*tag= m
FT FT /note= "rho-independent termination site"
FT FT 7695
FT FT /*tag= j
FT FT /label= ORF 3

```

FT EP635572-A2.

PN XX

XX XX

PD 25-JAN-1995.

XX 13-JUN-1994; 94EP-00108998.

XX 25-JUN-1993; 93US-00084709.

PR 06-MAY-1994; 94US-00239430.

XX (HOFF) HOFFMANN LA ROCHE &amp; CO AG F.

XX Bower SG, Perkins JB, Pero JG, Yocum RR;

XX WPI; 1995-053684/08.

XX Biotin genes, and constructs derived from *Bacillus subtilis* - for

XX improved production of recombinant biotin or biotin precursor for use in

XX e.g. dietary supplements.

XX Example II; Fig 14; 75pp; English.

XX AAQ81792 is the *B. subtilis* biotin operon and flanking sequences, as part

XX of an expression vector it can be used in the recombinant production of

XX biotin (or biotin precursor protein). The biotin can be used as a dietary

XX additive in animal feeds, and as a vitamin supplement for human

XX consumption. Biotin is also useful as a reagent for research, and

XX diagnostic procedures. (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 8478 BP; 2432 A; 1748 C; 2088 G; 2205 T; 0 U; 5 Other;

XX Query Match 11.7%; Score 139.2; DB 2; Length 8478;

XX Best Local Similarity 50.4%; Pred. No. 2e-33;

XX Matches 426; Conservative 0; Mismatches 408; Indels 12; Gaps 3;

QY 252 GACCGCGCGAGCATACAAAATCCGTTGATCGTGAACAAAGCCTTATCTCGCGGT 311

Db 5735 GAACGACCTGATCATAGACGATTGCGGCGCTTGCAGCGGAGCGTTTACGCGAGAAC 5794

QY 312 GATGAACCAATGGGACCGAGAAATTCAGAGAAATCACAGATGAACCTGATTCAAAATTCA 371

Db 5795 GACAGAGATTTCAGCGGATATCATTTGAAACCTGTCATTTGTTGATCAAGTGCA 5854

QY 372 GGGGCGCAGTGTGACCTTGTTCAGATTTTTCATACCCCGCTTCGGTTATGTGAT 431

Db 5855 AGGTAAAAAAGATGAGGCTCATTTGGAGCTTTGCTTTCTTTAGCAAGTTTGTGAT 5914

QY 432 ATCTGAGCTCTGGAGTGCCTTCAGCGGAGATGAACAGCTTTAAAGCATGGTCTGATCT 491

Db 5915 AGCTACATTATAGGTGTACCGGAGGAAGATAGGAGCAATTAAGGAGTGGGTGCGAG 5974

QY 492 TCTGGTCACTACACCGAGGATAAAGTGAAGAGCTGAAAAGCCTTTTGGAGAACG 551

Db 5975 TCTCAT-----TCAAACGATTGATTTTACCCGCTCAAGAAAGCAATTAAACAGAGGCA 6027

538	QY	TTTTTGGAAAGAACGAGATAGTGTGAGGAGAACTGGCCGCGTTTTTTTCGGCGCATCAT	597
577	Db	-----AGAGGAACTGCGGGCTTACATGCGGGGGTTGATC	612
598	QY	GAGAAAGAGGAAACAAACCGGAACAGGATATTATTTCTATTTTAGTGGAGCGGAAGAA	657
613	Db		672
658	QY	ACAGGCGAGAGCTGTTCGGTGAAGACTGATTCCTGTTTTCGACGCTGCTGCTGGTGGCC	717
673	Db		732
718	QY	GGAAATGAAACCACTACAAACCTGATTTCAAATGCGATGTACAGACATATTAGAAACGCCA	777
733	Db		792
778	QY	GGCGTTTACAGAGAACTGCGGAGCCATCTCTGAATGATGCTCAGGCAGTGGAGGAAGCC	837
793	Db		852
838	QY	TTGCGTTTCAGAGCGCGGCCCGCGTT-----TTGAGGCGCATTCGCAAGCGGGAACG	891
853	Db		912
892	QY	GAGATCGGGGGCACCTGATTTAAAGAAAGGTGATATGTTTGTGCGTTTGTGGCATCGGCA	951
913	Db		972
952	QY	AATCGTGATGAAGCAAGTTTGAAGACCGACATGTTTGATATCCGCGCCATCCCAAT	1011
973	Db		1032

```

RESULT 10
ACF06128
ID ACF06128 standard; DNA; 1215 BP.
XX
XX ACF06128;
XX
XX 30-SEP-2003 (first entry)
XX
XX Bacterial P450 enzyme encoding DNA SEQ ID NO:39.
XX
XX Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
XX KW diol; alkene; chirality; thermotolerance; thermostability; gene; ds.
XX
XX Eubacteria.
XX
XX Key Location/Qualifiers
FH 1. .1215
FT FT /*tag= a
FT FT /product= "P450 enzyme"
XX
XX WO2003052050-A2.
XX
XX 26-JUN-2003.
XX
XX 05-AUG-2002; 2002WO-US024910.
XX
XX 03-AUG-2001; 2001US-0309497P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Weiner D. Burke M. Hitchman T. Puiol C. Richardson T. Short J;
XX

```

1012 CGCATATTGGGTTGGCCACCGCATCCCATTTTGGCTTGGGCCCCCGCTTTC 1071  
 1033 CAGCACTCGGGTTCGGCCATGGCGTCCACCACTGCCTCGGGGGCGCGCTG 1092

1072 GA 1073  
 1093 GA 1094

RESULT 11  
 AAI99682\_39/c  
 Continuation (40 of 45) of AAI99682 from base 3900001 (Mycobacterium tuberculosis strain  
 WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

Fragment Name	Begin	End
AAI99682_00	1	110000
AAI99682_01	100001	210000
AAI99682_02	200001	310000
AAI99682_03	300001	410000
AAI99682_04	400001	510000
AAI99682_05	500001	610000
AAI99682_06	600001	710000
AAI99682_07	700001	810000
AAI99682_08	800001	910000
AAI99682_09	900001	1010000
AAI99682_10	1000001	1110000
AAI99682_11	1100001	1210000
AAI99682_12	1200001	1310000
AAI99682_13	1300001	1410000
AAI99682_14	1400001	1510000
AAI99682_15	1500001	1610000
AAI99682_16	1600001	1710000
AAI99682_17	1700001	1810000
AAI99682_18	1800001	1910000
AAI99682_19	1900001	2010000
AAI99682_20	2000001	2110000
AAI99682_21	2100001	2210000
AAI99682_22	2200001	2310000
AAI99682_23	2300001	2410000
AAI99682_24	2400001	2510000
AAI99682_25	2500001	2610000
AAI99682_26	2600001	2710000
AAI99682_27	2700001	2810000
AAI99682_28	2800001	2910000
AAI99682_29	2900001	3010000
AAI99682_30	3000001	3110000
AAI99682_31	3100001	3210000
AAI99682_32	3200001	3310000
AAI99682_33	3300001	3410000
AAI99682_34	3400001	3510000
AAI99682_35	3500001	3610000
AAI99682_36	3600001	3710000
AAI99682_37	3700001	3810000
AAI99682_38	3800001	3910000
AAI99682_39	3900001	4010000
AAI99682_40	4000001	4110000
AAI99682_41	4100001	4210000
AAI99682_42	4200001	4310000
AAI99682_43	4300001	4410000
AAI99682_44	4400001	4411529

Query Match 9.9%; Score 118.2; DB 4; Length 110000;  
 Best Local Similarity 48.1%; Pred. No. 5.9e-26;  
 Matches 401; Conservative 0; Mismatches 423; Indels 3; Gaps 2;

241 ATCAATTAACATGACCCCGCGAGCATACAAAATCCGTTTCAGTCGTGAACAAGCCCTT 300  
 55295 ATGATCGATGACATCCCGATCCCGACATCTGTTCCGGCGCAAGCTGTTAAGCGCGCTTC 55236

301 ACTCCCGCGGTGATGAAGCAATGGGACCGAGAAATTCAGAAATCAAGATGAATGATT 360  
 55235 ACCCGCAAGCGGTGAAGGACGAGGCGGTGATTCGCGCGCTGTTGTGACACCGCTGATC 55176

361 CAAAATTTTCGGGCGCAGTGAGTTTGACCTTGTTTTCAGATTTTTCATACCCGCTCCG 420  
 55175 GACGCGCTGTGGAACGCGCGGAGTGAGCTTCTGTCGGGACCTGCGCGCGCGCTACCG 55116  
 421 GTTATTGTGATATCTGAGCTGCTGGGAGTGCCCTTTCAGCGCAGATGGAACAGTTTAAAGCA 480  
 55115 ATGGCGGTGATCGGCGACATGCTCGGGTGCGTCCAGAGCAGCGGACATGTTCTTGGCG 55056  
 481 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGCTTTT 540  
 55055 TGGTCCGACGATCTGCTGA-CATTCCTCAGTTCGATGTTCTCAAGAGGATTTCCAGAT 54997  
 541 TTGGRAGACGAGATAAGTGTGAGGAGAACTGCGCGCTTTTTCGCGCATCATAGAA 600  
 54996 CACCATGAGCGCTTCGCGCCTTACACGACTTCACCGG-----GCCACCATTGGC 54945  
 601 GAAAACGGAACAAACCCGGAACAGGATATTATTTCATTATTAGTGAAGCGGAAGAAACA 660  
 54944 GCACGCGGAGCGGACCCACCGACGACCTGGTCAGCGTGTGCTGGTTCGGAAGTTGAC 54885  
 661 GCGGAGAAGCTGTCGGTGAAGAGCTGATTCGTTTGGACGCTGCTGCTGGTGGCGGA 720  
 54884 GCGAGCGGCTAAGCGACGACGAGCTGGTCANGAGACGCTGCTGATCCTGATCGGCGGC 54825  
 721 AATGAAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAGAAAACGCCAGGC 780  
 54824 GACGAGACACGCGGATACCTTGAGCGGTGGTACCGAGCAGCTGTGCGCAACCGTGAC 54765  
 781 GTTTACGAGAACTGCGGACCCATCCTGAACACTGATGCTCAGCAGGTGGAGGAGCCTTG 840  
 54764 CAGTGGGACCTGCTGACGCGGACCCGCTGCTGCTGCGCGGCGCATCGAGAGATGCTA 54705  
 841 CGTTTCAGAGCGCGCGCGCTTTTGAAGCGCATTCGCAAGCGGATACGAGATCGGG 900  
 54704 CPTTGAACCGCGCGCTTAAAGAAACATGTGCGCGGTGTGACCGCGGATACCGAGTTTCAC 54645  
 901 GGGCACTGATTAAGAAGTGATATGTTTTCGCGCTTTCGCGCATCGGCAATTCGTGAT 960  
 54644 GGCACGCGGTGTGTCGCGCGGAGAGATGATGCTCTCTCGAGTGGCGGAACTTCGAC 54585  
 961 GAACAAAGTTTGCACAGACCGCACATGTTTGTATATCGCGCGCATCCCAATCCGCATATT 1020  
 54584 GAGCGGTTTCTGTGAACCGGAAAGTTTGATGTTTCAGCGAAATCCAAACAGCCACTTG 54525  
 1021 GCGTTGGCACCGCATCTCAATTTTTCCTTGGGCGCGCGCTTCCCGCTCTGA 1073  
 54524 GCGTTTGGCTTCGGCACGCAATTTCTGCTGGCAATCAGCTGGCCCGGTTGGA 54472

RESULT 12  
 AAI99683\_39/c  
 Continuation (40 of 44) of AAI99683 from base 3900001 (Mycobacterium tuberculosis strain  
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

Fragment Name	Begin	End
AAI99683_00	1	110000
AAI99683_01	100001	210000
AAI99683_02	200001	310000
AAI99683_03	300001	410000
AAI99683_04	400001	510000
AAI99683_05	500001	610000
AAI99683_06	600001	710000
AAI99683_07	700001	810000
AAI99683_08	800001	910000
AAI99683_09	900001	1010000
AAI99683_10	1000001	1110000
AAI99683_11	1100001	1210000
AAI99683_12	1200001	1310000
AAI99683_13	1300001	1410000
AAI99683_14	1400001	1510000
AAI99683_15	1500001	1610000
AAI99683_16	1600001	1710000
AAI99683_17	1700001	1810000
AAI99683_18	1800001	1910000



Query Match 9.8%; Score 116.4; DB 4; Length 110000;  
Best Local Similarity 55.4%; Pred. No. 2.3e-25;  
Matches 225; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 669 GCTGTCCGGTGAAGAGCTGATTCGCTTTTGACGCTGCTGTGTGGCGGAAATGAAC 728  
DB 72139 GCTGTCCGACCGCACTGTACTCTTCTTCCACTGTACTGTTCAGCGCGCGCGGAAAC 72198

QY 901 GGCACCTGATTAAGAGGTGATGATGTTTGGCGTTTGGCATCGGCAATCGTAT 960  
DB 47374 GGCACGCGTGTGTGCGCGGAGAGATGATGCTGCTCTTCGAGTCGGGAACTTCGAC 47315  
QY 961 GAAGCAAGTTTCACAGACGCGCATGTTTGATATCGCGCCATCCCAATCCGCATAT 1020  
DB 47314 GAGCGGTTTCTGTGAACCGGAAAGTTTGATGTTTCAGCGAAATCCAAACAGCCACTTG 47255  
QY 1021 GCGTTTGGCCACGCGCATCTTTTTCCTTTGGGCGCCCGCTTGCCTTCTTGA 1073  
DB 47254 GCGTTTGGCTTCGCGACGATTTCTGCTCGCAATCAGTCGCGCGTGGGA 47202

## RESULT 13

AAI99682\_09  
Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H:  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match

Score 116.4; DB 4; Length 110000;

Best Local Similarity 55.4%; Pred. No. 2.3e-25;

Matches 225; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 669 GCTGTCCGGTGAAGAGCTGATTCGCTTTTGACGCTGCTGTGTGGCGGAAATGAAC 728

DB 72139 GCTGTCCGACCGCACTGTACTCTTCTTCCACTGTACTGTTCAGCGCGCGCGGAAAC 72198

QY 729 CACTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAGAAACCCAGCGGTTTACGA 788  
DB 72199 CACCCGTAACCTCAATTCGCGGGGCTGCTGGCGTGGCGGAGAACCTTGACCACTGCA 72258  
QY 789 GGAACCTGGCGAGCCATCCCTGAAGTGTCTCAGGAGTGGAGAGCTTGGGTTTCAG 848  
DB 72259 AACGCTGCGAAGCGATTTTGAGTTGTTCGCTGCGATCGAAGAGATCGTAGAGTGGAC 72318  
QY 849 AGCGCGCGCCCGGTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGGGGGCACT 908  
DB 72319 GTCGCGCTCACCATCGAACGCGGACGCGGTCCCTGCGGTGAGCTGGCGGCGAGCC 72378  
QY 909 GATTAAAGAGTGATATGTTTGGCTTTGCGCATCGGCAATTCGATGATGACGAA 968  
DB 72379 GATCGAGCGGGTCAAGAGTTGTGTGGAGAGGCTCGGCGCAACCTGATCCAGCGT 72438  
QY 969 GTTTCAGACGCGCATGTTTGATATCGCGCCATCCCAATCCGCATATTCGGTTTG 1028  
DB 72439 GTTCGACCGCGGAGAGTTGATATCACCGAAACCAATCCGACCTGGGTTTCGG 72498  
QY 1029 CCACGCGATCCATTTTTCGCTTGGGCGCGCTTGGCGGCTTCTTGAA 1074  
DB 72499 TCAGGGGTGCACTATTGCTGGCGCCATCTGGCTCGGCTGAA 72544

## RESULT 14

AAI99683\_08  
Continuation (9 of 44) of AAI99683 from base 800001 (Mycobacterium tuberculosis strain H37Rv)  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

Fragment Name	Begin	End
WP AAI99683_00	1	110000
WP AAI99683_01	100001	210000
WP AAI99683_02	200001	310000
WP AAI99683_03	300001	410000
WP AAI99683_04	400001	510000
WP AAI99683_05	500001	610000
WP AAI99683_06	600001	710000
WP AAI99683_07	700001	810000
WP AAI99683_08	800001	910000
WP AAI99683_09	900001	1010000
WP AAI99683_10	1000001	1110000
WP AAI99683_11	1100001	1210000
WP AAI99683_12	1200001	1310000
WP AAI99683_13	1300001	1410000
WP AAI99683_14	1400001	1510000
WP AAI99683_15	1500001	1610000
WP AAI99683_16	1600001	1710000
WP AAI99683_17	1700001	1810000
WP AAI99683_18	1800001	1910000
WP AAI99683_19	1900001	2010000
WP AAI99683_20	2000001	2110000
WP AAI99683_21	2100001	2210000
WP AAI99683_22	2200001	2310000
WP AAI99683_23	2300001	2410000
WP AAI99683_24	2400001	2510000
WP AAI99683_25	2500001	2610000
WP AAI99683_26	2600001	2710000
WP AAI99683_27	2700001	2810000
WP AAI99683_28	2800001	2910000
WP AAI99683_29	2900001	3010000
WP AAI99683_30	3000001	3110000
WP AAI99683_31	3100001	3210000
WP AAI99683_32	3200001	3310000
WP AAI99683_33	3300001	3410000
WP AAI99683_34	3400001	3510000
WP AAI99683_35	3500001	3610000
WP AAI99683_36	3600001	3710000
WP AAI99683_37	3700001	3810000
WP AAI99683_38	3800001	3910000
WP AAI99683_39	3900001	4010000
WP AAI99683_40	4000001	4110000
WP AAI99683_41	4100001	4210000
WP AAI99683_42	4200001	4310000

WP	AAI99683_43	4300001	4403765
Query Match	9.8%	Score 116.4;	DB 4; Length 110000;
Best Local Similarity	55.4%	Pred. No. 2.3e-25;	
Mismatches	225;	Conservative	0; Mismatches 181; Indels 0; Gaps 0;
QY	669	GCTGTCGGGTGAGAGCTGATTCGGTTTTCAGCGTCTGCTGGTGGCGGAAATGAAC	728
DB	74269	GCTGTCGGACGCGGAATGTACCTGTCTTCATCTACTTTCAGCGCGCGGGAAC	74328
QY	729	CACTACAAACCTGATTTCAAAATCGGATGTACAGCATATTAGAAACGCGGCGCTTACGA	788
DB	74329	CACCCGTAACCTCAATTCGCGCGGCTGCTGGCGTGGCGAGAACCTTGACCACTGCA	74388
QY	789	GGAACCTGGCGAGCCATCCCTGMACTGATGCTCAGGCGAGTGGAGAGCCTTGGTTTCAG	848
DB	74389	AACGCTGCGAAGCGATTTTCAAGTTGTTCGCACTGCGATCGAAGAGATCTGAGGTTGGAC	74448
QY	849	AGCGCGCGCCCGGTTTTCAGGCGCATTCGCAAGCGGACACGAGATCGGGGGGCACT	908
DB	74449	GTGCGGTGACCATCGAAGCGGCGACGGGTCTCGGTCGCTGCGGTGAGCTGGCGGCGAGCC	74508
QY	909	GATTAAAGAGTGATATGCTTTTGGGTTTGGCATCGGCAATCCCAATCCGCATATTCGGTTGG	968
DB	74509	GATCGAGCGGCTCAGAAGGTTGTGGTGGAGGGCTCGGCGCAACCGTGATCCAGCGT	74568
QY	969	GTTCGACAGCCGACATGTTTGATATCCGCGCGCATCCCAATCCGCATATTCGGTTGG	1028
DB	74569	GTTCGACCGCGGAGGTTGATATCACCGAAGAACCAATCCGACCTGGGTTTCGG	74628
QY	1029	CCACGCGATCCATTTTTCGCTTGGGCGCGCTTGGCGGCTTCTTGAA	1074
DB	74629	TCAGGGGTGCACTATTGCTGGCGCCATCTGGCTCGGCTGAA	74674
RESULT 15			
ACF06136	ID	ACF06136 standard; DNA; 1248 BP.	
XX	AC	ACF06136;	
XX	DT	30-SEP-2003 (first entry)	
XX	DE	Bacterial P450 enzyme encoding DNA SEQ ID NO:55.	
XX	XX	Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;	
XX	XX	diol; alkene; chirality; thermostability; gene; ds.	
XX	OS	Eubacteria.	
XX	Key	Location/Qualifiers	
XX	CDS	1..1248	
FT		/*tag= a	
FT		/product= "P450 enzyme"	
XX	PN	W02003052050-A2.	
XX	XX	26-JUN-2003.	
XX	XX	05-AUG-2002; 2002WO-US024910.	
XX	PF	03-AUG-2001; 2001US-0309497P.	
XX	PR	(DIVE-) DIVERSA CORP.	
XX	PA	Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short C;	
XX	PI	WPI; 2003-541641/51.	
XX	DR	P-PSDB; ABR82146.	
XX	XX	Novel polypeptide having P450 enzyme activity and polynucleotides	
XX	XX	encoding the polypeptide, useful for catalyzing the hydrolysis of	
XX	XX	epoxides and arene oxides to their corresponding diols.	

XX PS Claim 18; Page 364; 365pp; English.

XX CC The present sequence encodes a bacterial p450 enzyme. p450 enzymes can be

XX CC used to catalyze the hydrolysis of epoxides and arene oxides to their

XX CC corresponding diols. p450 enzymes can also be used for hydrolysing an

XX CC alkene, for producing a compound of a desired chirality, and for

XX CC increasing thermostolerance or thermostability of a p450 polypeptide

XX SQ Sequence 1248 BP; 127 A; 451 C; 510 G; 160 T; 0 U; 0 Other;

Query Match 9.6%; Score 114.2; DB 8; Length 1248;

Best Local Similarity 46.7%; Pred. No. 8.6e-26;

Matches 399; Conservative 0; Mismatches 453; Indels 3; Gaps 1;

QY 230 TTGGAATTCATCAATCAATGATGAGCAATGGGAACGAGAAATTCAGAAATCCGTTCAAGTCGTGA 289

DB 263 TCAGAACTGGCTGCTTCTCTGACCCGCGCCGACACGCGGCTGCGGCCGAGATCGCG 322

QY 290 ACAAGCCTTTACTCCGCGGTGATGAAGCAATGGGAACGAGAAATTCAGAAATTCACAG 349

DB 323 CCCCGCGGTGAGCGCTCGGCGGTACGCGCTTGGTCCCGGTACGCGGATCGCG 382

QY 350 ATGAACGTATCAAAATTCAGCGGCGAGTGAGTTGACCTTGTTCAGATTTTCAT 409

DB 383 AGGAATCTGTACGCGCCCTCGCGCGCGCCGCGTGGTCAAGTGGTCGAGGGCTTCGCG 442

QY 410 ACCCGCTTCGCGTTATGTGATATCTGAGCTGTGGAGTGCCCTTCAGCGCAGATGGAAC 469

DB 443 CCCCGTTCCGCTGTGCTGCTGCGGCGCTGCTCGGGTGCACCCGCTCGGTGGCGGT 502

QY 470 AGTTAAAGCATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAGTGAAGAGCTG 529

DB 503 GGTTCGCGGAGGAGGCG---CTGCGCTCCAGCGGTCCGCGGACCCAGCGGCGACCGGT 559

QY 530 AAAAGCCTTTTGGAGAACGAGATGAGTGTGAGGAGAACTGGCGCGCTTTTTCGCG 589

DB 560 CGCGCGCGCGCTCGCACGCGGCGCGACCGCGCGCGCGACCTTGACGCGTACTTCGCGG 619

QY 590 GCATCATAGAGAAAGCGAAACAAACCGGAACAGAGATATTATTTCTATTATTAGTGAAG 649

DB 620 CGAGCTGGCGCGCGCGCTCGAGGACCGGGGGACCTGCTCTCGCGCTGCGCGG 679

QY 650 CGGAAGAAACAGCGGAGAACTGTGCGGTGAAGAGCTGATTCGTTTTGACGCTGCTGC 709

DB 680 CGCGCGCGGAGACCCCTGCTGGGAGACGACCGCGCTGACCTGACCTGCTCCACCTCC 739

QY 710 TGGTGGCGGAATGAACCACTACAACTGATTTCAATGCGATGTACAGCATATTAG 769

DB 740 TGACGCGCGGCGAGGACGACGCGGCGCTGCTGGGAGGCGGCTGCTGCGCTGCTGG 799

QY 770 AAACGCGCGGCTTTACAGGAACTGGCGAGCCATCTGAACTGATGCTCAGGAGTGG 829

DB 800 CGCGCGCGAGTGGCGGAGGAACTGCGCGGACCCCGGCTGTTGCCGAAACGCGTGG 859

QY 830 AGGAGCCTTTGCGTTTCAGAGCGCGCGCGCGTTTGGAGCGCATTCGCAAGCGGATA 889

DB 860 ACAGATTCTCTGCGCACGACCCACCGTGCAGATGGTCAACGCGTGGCGCGGAGCG 919

QY 890 CGGAGATCGGGGGCACCTGATTAAGAGGTGATATGTTTGGCGTTTGTGGCATCGG 949

DB 920 CGGAGCTGGCGCGCGAGCGCTCGCGGGGCGACCGGCTCCAGCTGCTGCTGGGTTCGG 979

QY 950 CAATCTGTGATGAAGCAAGTTTGACAGACCGCACATGTTTGATATCGCGCGCATCCCA 1009

DB 980 CCCACCGGACCCCGCGCTTCCGCGACCCCGACCGGCTGGACATCCGCGGGACACCG 1039

QY 1010 ATCCGCATATTGGTTGGCCACCGGCAATTCATTTTGTGCTTGGGCGCGCGCTGCGCGTC 1069

DB 1040 GCCCGCATGCGCGTTCGGTCTGCGCATCCACTACTGCTGGGCGCGCGCTGCGCACGCG 1099

QY 1070 TTGAAGCAATATCG 1084

DB 1100 CGGAGCGCGGAGATCG 1114

Search completed: May 28, 2004, 20:33:07

Job time : 406 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 20:08:15 ; Search time 2408 Seconds  
(without alignments)  
14769.874 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgtaaacgcgcg.....gcttcgtgtgaaatgtaa 1191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_esthum:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106.2	8.9	975	28	BZ675691 PUBDH78TD
2	81.2	6.8	1285	28	BZ567880 pacs2-164
3	77.4	6.5	479	14	CD091452
4	72.6	6.1	1127	28	BZ548452 pacs1-60

5	53.6	4.5	330	28	AV174775
6	53.4	4.5	636	14	CD486042
7	52	4.4	1223	28	BZ551987
8	49.4	4.1	638	14	CD486072
9	44.2	3.7	1101	29	CNS0039G
10	43.4	3.6	588	28	AZ934336
11	42.8	3.6	1136	28	BZ563995
12	42.8	3.6	1201	13	BX381961
13	42.2	3.5	718	29	CG092084
14	42.2	3.5	432	13	BX577443
15	40.6	3.4	1201	13	BX406481
16	40.6	3.4	1201	13	BX417669
17	40.6	3.4	1201	13	BX442405
18	40	3.4	244	14	CF050866
19	40	3.4	2096	11	AY109482
20	39.8	3.3	591	14	CB923439
21	39.6	3.3	601	28	BZ374029
22	39.4	3.3	827	29	CG041562
23	39.4	3.3	1201	13	BX361080
24	39.4	3.3	1201	13	BX376097
25	39.2	3.3	575	12	BG515155
26	39.2	3.3	654	9	AL888423
27	39.2	3.3	658	9	AL862443
28	39.2	3.3	671	9	AL886494
29	39	3.3	1091	13	BX424950
30	38.8	3.3	469	9	AA246055
31	38.8	3.3	636	14	CD429348
32	38.8	3.3	641	12	BM329835
33	38.6	3.2	952	29	CNS0605L
34	38.6	3.2	1023	28	CC292361
35	38.6	3.2	1057	28	CC293665
36	38.4	3.2	247	14	CK228855
37	38.4	3.2	268	14	CK228977
38	38.4	3.2	306	14	CK218284
39	38.4	3.2	503	14	CK228926
40	38.4	3.2	523	14	CK228927
41	38.4	3.2	552	14	CK228929
42	38.4	3.2	563	14	CK228925
43	38.4	3.2	593	9	AW111487
44	38.4	3.2	595	14	CK228997
45	38.4	3.2	624	14	CK228974

#### ALIGNMENTS

RESULT 1  
BZ675691 PUBDH78TD ZM\_0.6 1.0 KB Zea mays genomic clone ZMMBTa029M12,  
LOCUS 975 bp DNA linear GSS 05-FEB-2003  
DEFINITION Genomic survey sequence.

ACCESSION BZ675691

VERSION BZ675691.1 GI:28226790

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE Rukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 975)

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and

Bennetzen, J.

Maize Genomics Consortium

Unpublished (2003)

Contact: Cathy Whitelaw

TIGR Medical Center Drive, Rockville, MD 20850, USA

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

## FEATURES

source  
1. .975  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="ZMMBTA029M12"  
/clone\_lib="ZM 0.6 1.0 KB"  
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
cot selected genomic DNA library"

## ORIGIN

Query Match 8.9%; Score 106.2; DB 28; Length 975;  
Best Local Similarity 51.8%; Pred. No. 4.2e-20;  
Matches 268; Conservative 0; Mismatches 243; Indels 6; Gaps 1;  
QY 564 GGAAGAACTGCGCCGCTTTTGGCGGCATCATAGAGAAAGCGAAGCAACACCGGACCA 623  
DB 157 GGGCGAGCTGGCCGATACATCAAAAACCTGGTGGCCGACGTGACGCGCAACCAACGCA 216  
QY 624 GATATTATTTCTATTTAGTGGAGCGGAAGAAACAGCGCAGAGCTGTCCGGTGAAGA 683  
DB 217 TGACTTCATGACCCCTGACCCAGCGCAGAAAGGCGAGTCCCTCAACGAGGACCA 276  
QY 684 GCTGATTCGGTTTTCAGCTGCTGCTGGTGGCGGAAATGAAACCACTACAAACCTGAT 743  
DB 277 ATTGCTGCTGTGTGTCATTCGATCTCTGCTGGTGGCGGTTACGAGCGCAGTGCCTGCAGAT 336  
QY 744 TCCATATGCATGTACAGCATATAGAAACCGCGAGGCTTTACAGGACCTGCGCAGCA 803  
DB 337 CCCCAACTTTATCCAGTGTCTGTGACAAACCCCGCAGTGGCAACAGCTCAAGAGCCGA 396  
QY 804 TCCTGAACATGCTCAGCGAGTGGAGAGCCCTTGGCTTTTCAGAGCGCGCGCCCGCGG- 862  
DB 397 TCCGAGCAGATACCGCAGCGGTTCGAGAGCTGTGCGCTACATCCGCTGGCTCGGC 456  
QY 863 -----TTTTCAGGCGCATTCGCAAGCGGATACGGAGATCGGGGGGACCTGATTAAGA 917  
DB 457 GGGGATGTTGTGCACTACGCGCTGGAAGACATTCAGTCTCGTGAACCCCTGTGGCGCA 516  
QY 918 AGTGATATGTTTGGCGTTTGGCATCGGCAATCGTATGATGATGATGATGATGATGATGAT 977  
DB 517 GGGCGAGCGGCTGTGCTGATGCGCGCGCGCAACATGACCCGCGCCCTTTGAAA 576  
QY 978 ACCGCAATGTTGATATCGCGCCCATCCCAATCCGATATTCGTTTGGCCACGCGAT 1037  
DB 577 CCCACAGGCGTGAACCTGCAAGCGGATGCCAGCGCCATTTGCGCTTTGGCCACGCGCT 636  
QY 1038 CCAATTTTGGCTTGGGCGCGCGCTTGGCGCTTGGAA 1074  
DB 637 GCATCACTGATCGGCTCAGCCCTGGCGCGGTCGAA 673

## RESULT 2

BZ567880 1285 bp DNA linear GSS 17-DEC-2002  
LOCUS pacs2-164.7256.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pacs2-164.7256, genomic survey sequence.

BZ567880.1 GI:27199936

## KEYWORDS

GSS.

## ORGANISM

Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1285)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

## source

1. .1285  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone\_lib="pacs2-164.7256"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun  
library."

## ORIGIN

Query Match 6.8%; Score 81.2; DB 28; Length 1285;  
Best Local Similarity 50.7%; Pred. No. 1.6e-12;  
Matches 222; Conservative 0; Mismatches 213; Indels 3; Gaps 1;  
QY 541 TAGTGGAGCGGAACACAGCGAGAGCTGTCCGTTGAAGAGCTGATTCGTTTTCGA 700  
DB 111 TGGTGCAGCGCGCGAGCAGAGCGGCCAGTTGAGCGAAGCGGAACCTGCTCCATGGGCC 170  
QY 701 CGCTGCTGCTGTGGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACA 760  
DB 171 ACCTGCTGATGATGAGCGGCTTCGAGACCACTGATCGGCAACCGCGTGTGCA 230  
QY 761 GCATATTAGAAACCGCAGCGCTTTTACAGAGAACTTCGCGAGCACTCTGAATGATGCTTC 820  
DB 231 CCCTGCTGTTCAACCGCGAGCACTGCGTTGCTGCGGGCGCAGCGGAACTCTCGCCA 290  
QY 821 AGGCAGTGGAGGAA--GCCTTGGCTTTTCAGAGCGCGCGCCCGCTTTTGAAGCGCATG 877  
DB 291 ACGCATGAGAAACTGTGTCGCCACGACATCCGTTCCGCGGCTCGGATGTTCCTTCA 350  
QY 878 CCAAGCGGATACGAGATCGGGGCGCACCTGATTAAGAAAGGTGATATGTTTTGGCGT 937  
DB 351 CCGTGGAGACGTGGAACCTGACCGGGTCACTATCCCGCGGGAATACATCTCGTCT 410  
QY 938 TTGTGGCATCGGCAATCGTATGAGCAAGATTGACAGACCGCATGTTGATATCC 997  
DB 411 CCAACCTGACCGCAACACCATGCGAGCGCTTCGACGATCCCGACCGCTCGACCTCA 470  
QY 998 GCCGCATCCCAATCCGATATTCGTTTGGCCAGCGCATCCATTTTTCCTTTGGGGGCC 1057  
DB 471 CCGCAACACGATGCGCATCTCGCTACGGCTTCGGGCTGCACTACTGCTGGCGGCT 530  
QY 1058 CGCTTGGCGGCTTTTGAAG 1075  
DB 531 CGCTGGCGCGGCTGGAGG 548

## RESULT 3

CD091452/c

## LOCUS

CD091452 479 bp mRNA linear EST 14-SEP-2003  
DEFINITION MCI-0087P-A156-F03-U.G MCI-0087 Schistosoma mansoni cDNA clone

MCI-0087P-A156-F03.G, mRNA sequence.

## ACCESSION

CD091452

## VERSION

CD091452.1 GI:34642252

## KEYWORDS

EST.

## SOURCE

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 479)

Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,

Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y., Jr.,

Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,

Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,

Leite, R.A., Malaquias, L.C.C., Marques, R.C.F., Miyasato, P.A.,

Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,  
Sa.R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,  
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,  
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.  
Transcriptome analysis of the acelomate human parasite Schistosoma  
mansoni

Nat. Genet. 35 (2), 148-157 (2003)  
22879926

Contact: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,  
Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

E-mail: verjovski@usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST  
Genome Project. All sequences in the project were assembled and  
annotated. This entry and all the assembled sequences can be seen  
in the following URL <http://bioinfo.ig.usp.br/schisto/>  
Plate: MCI-0087P-A156 row: 3 column: F.

## FEATURES

source

1. 479  
/organism="Schistosoma mansoni"  
/mol\_type="mRNA"  
/db\_xref="taxon:6183"  
/clone="MCI-0087P-A156-F03.G"  
/sex="mixed pool"  
/dev\_stage="Cercaria"  
/lab\_host="Biomphalaria glabrata"  
/clone\_lib="MCI-0087"  
/note="Vector: pGEM T-easy"

## ORIGIN

Query Match 6.5%; Score 77.4; DB 14; Length 479;  
Best Local Similarity 49.9%; Pred. No. 1.4e-11;  
Matches 223; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
QY 334 ATTCAGAAATCACAGATGAACCTGATTTAAAGATTTTACGGGGCGAGTGTGACCTT 393  
Db |||||  
QY 456 ATTCAGAGATGCTGAGCGCTGATTGATTAATGACAGTGGCTAAGGGCGAATGGAATT 397  
Db |||||  
QY 394 GTTCAGATTTTCATACCGCTTCGGTTATGTGATATCTGAGTGTGGAGTGCCT 453  
Db |||||  
QY 396 ATTCGGAATTCGCCCTTCTGTTCCGATTACAGTAATTTGTGAATGCTGGGATTAACCT 337  
QY 454 TCAGGCGAGATCGAAACAGTTTAAAGCATGCTGATCTTCTGTCAGTACACCGAAGGAT 513  
Db |||||  
QY 336 TCGGCTGACCAGAGAAATTCAGACAAATGCTCAACGCAATGACTCAGACCTTAGAGCT 277  
QY 514 AAAAGTGAAGAGCTGAAGAAAGCTTTTGGAGAAACGAGATAAGTGTGGAGAAAGCTG 573  
Db |||||  
QY 276 GSCAGCATCACAAATCTTCAGTCTTAC---ACGTATCAAAATCACGCTGGTGAAGCTG 220  
QY 574 GCCGGTTTTTTCGGGCATCATAGAAAGAGCGAAACCGAACCAGCAACGATATATT 633  
Db |||||  
QY 219 CATGACTATCTCGAACCGCTAGTGCAGGAGCGCGAAAGAAATCCAAAGCGGATCTAAT 160  
QY 634 TCTATTTAGTGAAGCGGAAGAAACAGCGAGAGCTGTCGGTGAAGAGCTGATTCGG 693  
Db |||||  
QY 159 ACAGCTCTGTTGAGCAGAAATGAGACAAATACTTGAAGATCAGAGTAATTGGA 100  
QY 694 TTTTGCACGCTGCTGCTGTGGCCGGAATGAACCACTACAACTGATTTCAATGCG 753  
Db |||||  
QY 99 AATATAATGCTTCTGCTAGTGCAGGCGCAGACAAACCGTTAATTTAATTGGCAACGGT 40  
QY 754 ATGTCAGCATATTAGAAAGCCGACGC 780  
Db |||||  
QY 39 ATGTTGCTTTATGCAACATCCGGAC 13

RESULT 4  
BZ548452

LOCUS BZ548452 1127 bp DNA linear GSS 17-DEC-2002  
DEFINITION pacsl-60.1124.s1 pacsl-60 Pseudomonas aeruginosa genomic clone  
pacsl-60.1124, genomic survey sequence.  
ACCESSION BZ548452  
VERSION BZ548452.1 GI:27152033  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 1127)  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

## FEATURES

source

1. 1127  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pacsl-60.1124"  
/clone\_lib="pacsl-60"  
/note="clinical isolate 1-60 Whole genomic shotgun  
library."

## ORIGIN

Query Match 6.1%; Score 72.6; DB 28; Length 1127;  
Best Local Similarity 57.1%; Pred. No. 5.7e-10;  
Matches 132; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 240 CATCTAATACATGACCGCGGAGCATACAAAATCCGTTGAGTGGTGAACAAAGCCTT 299  
Db |||||  
QY 250 CATGCTCACTCGACCGCGGACCATACCCGCTCGCTGGTTGGCGCGGCTT 309  
QY 300 TACTTCGCGCTGATGAAGCAATGGGAAACGAGAAATCAAGAAATACAGATGACTGAT 359  
Db |||||  
QY 310 CACCCCGCGCTGAGCGCTGCAACCGCATATACACGATCACCGAGGAATTGCT 369  
QY 360 TCANAATTCAGGGCGGAGTGTGACCTTGTTCAGATTTTTCATACCGGCTTC 419  
Db |||||  
QY 370 GGACGCTGCGCGCGGCAACAGCGGACCTGTGATGGCGACTTCGCGATCCCGCTGAC 429  
QY 420 GGTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACA 470  
Db |||||  
QY 430 CATCGCGGTGATCTTCGAGCTGCTGGGCTATCCCGAGCGGCGGCGGAACA 480  
RESULT 5  
AY174775 330 bp DNA linear GSS 13-JAN-2003  
LOCUS AY174775  
DEFINITION Mycobacterium avium subsp. paratuberculosis DNA  
Mycobacterium avium subsp. paratuberculosis genomic clone  
MtbH1.15, genomic survey sequence.  
ACCESSION AY174775  
VERSION AY174775.1 GI:27689314  
KEYWORDS GSS.  
SOURCE Mycobacterium avium subsp. paratuberculosis  
ORGANISM Mycobacterium avium subsp. Paratuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
avium complex (MAC).  
REFERENCE 1 (bases 1 to 330)  
AUTHORS Nielsen,K.K. and Ahrens,P.

**TITLE** Subtractive hybridization and bioinformatics: combining genetic methods to obtain new information on species-specific elements of Mycobacterium avium subspecies paratuberculosis

**JOURNAL COMMENT**

Unpublished (2003)  
Contact: Nielsen KK  
Bacteriology  
Danish Veterinary Institute  
Bulowsvej 27, Copenhagen V, 1790, Denmark  
Email: kni@veti.dk  
Class: unknown.

**FEATURES** source  
1..330 Location/Qualifiers  
/organism="Mycobacterium avium subsp. paratuberculosis"  
/mol\_type="genomic DNA"  
/sub\_species="paratuberculosis"  
/db\_xref="taxon:1770"  
/clone="MPtb61.15"  
/clone\_lib="Mycobacterium avium subsp. paratuberculosis DNA"

**ORIGIN**

Query Match 4.5%; Score 53.6; DB 28; Length 330;  
Best Local Similarity 48.7%; Pred. No. 0.00017;  
Matches 146; Conservative 0; Mismatches 154; Indels 0; Gaps 0;  
  
604 AAGCGAAACAAACCGAACAGATATTTCTATTTAGTGAAGCGGAGAAACAGGC 663  
|||||  
26 AAGCGGCGGAGCCAGGTGACGATATCGCACTTCTTACTCCAGCGCGAGGTGATGGT 85  
|||||  
664 GAGAGCTGTCGGTGAGAGCTGATTCGGTTTCACGCTGCTCTGGTGCCCGAAT 723  
|||||  
86 CAGAGCCTCAGCGATTTGGAAATTTAACTGTCTTATGTTGCTGATCGATCGCGAGT 145  
|||||  
724 GAAACCACTACAAACCTGATTTCAAATCGGATGTACAGCATATTAGAAACGCCAGCGGT 783  
|||||  
146 GATAGACCCCAATCTGTAGACGGGAATCTTGCTCTGCTTGACATCCCCAAGAA 205  
|||||  
784 TACAGGAACCTGGCAGCCATCTCGAATGATGCTCAGGAGTGGAGAACCTTGCGT 843  
|||||  
206 CTGACGCGGCTTAAGGCAGACCCCTTCACTATCGCGACCGGATCGAAGAAATCTGGGC 265  
|||||  
844 TTCAGAGCGCGCCCGGTTTGGAGCGCATTCGACAGCGGATACGAGATCGGGGG 903  
|||||  
266 TACACGTCGCGGTGACGGCATCTCGCGCACCGCGCAAGACACTGAGTTGCGCGGG 325  
|||||

**RESULT 6**  
CD486042 636 bp mRNA linear EST 01-JUL-2003  
LOCUS CFUS5.3A03 Cotton Root and Hypocotyl tissues infected with Fusarium  
DEFINITION oysporium f.sp. vasinfectum (Fov) Lambda ZIPLOX Library (CFUS)  
Gossypium hirsutum cDNA clone CFUS5.3A03 5' similar to Chain A,  
X-Ray Structure Of Nitric Oxide Reductase, mRNA sequence.

**ACCESSION** CD486042  
**VERSION** CD486042.1 GI:31407007  
**KEYWORDS** EST.  
**SOURCE** Gossypium hirsutum (upland cotton)  
**ORGANISM** Gossypium hirsutum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.  
1 (bases 1 to 636)

**REFERENCE** DOWD, C., WILSON, I. and McFadden, H.  
**AUTHORS** Different Gene Expression Responses in Cotton Root and Hypocotyl  
**TITLE** tissues during infection with Fusarium Wilt Disease  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Cairtriona Dowd, Helen McFadden  
Commonwealth Scientific and Industrial Research Organisation  
Division of Plant Industry  
Black Mountain Laboratories, Cnr Clunies Ross Street & Barry Drive,  
Black Mountain, Canberra, ACT, 2601, Australia  
Tel: 61 2 6246 4914, 6246 5377  
Fax: 61 2 6246 5000

Email: Cairtriona.Dowd@csiro.au, Helen.McFadden@csiro.au  
Vector clipped sequences Bases 1-17 (GTGACCCACGCTCCG): Sali  
adapter  
Seq primer: M13 reverse primer  
High quality sequence stop: 636.  
Location/Qualifiers

**FEATURES** source

1..636  
/organism="Gossypium hirsutum"  
/mol\_type="mRNA"  
/cultivar="DeltaEMERALD"  
/db\_xref="taxon:3635"  
/clone="CFUS5.3A03"  
/tissue\_type="Fov infected root and hypocotyls tissues"  
/dev\_stage="seedlings, pooled from 5 hours to 7 days post  
infection with Fov."  
/lab\_host="V1090(ZL)"  
/clone\_lib="Cotton Root and Hypocotyl tissues infected  
with Fusarium oysporium f.sp. vasinfectum (Fov) Lambda  
ZIPLOX Library (CFUS)"  
/note="Vector: Lambda ZIPLOX; Site 1: Sali; Site 2: NotI;  
mRNA was prepared from pooled root and hypocotyl tissues  
of the cotton cultivar DeltaEMERALD at different stages of  
Fov infection from 5 hours to 7 days post infection. cDNA  
was synthesised from a NotI-oligo dT primer/adaptor using  
the manufacturers protocols (Life Technologies) and then  
ligated to a Sali adapter to facilitate directional  
cloning. The cDNA was cloned into the Sali and NotI sites  
of the Lambda ZIPLOX phage vector (Life Technologies).  
Constructed by Cairtriona Dowd and Helen McFadden."

**ORIGIN**

Query Match 4.5%; Score 53.4; DB 14; Length 636;  
Best Local Similarity 49.0%; Pred. No. 0.00027;  
Matches 172; Conservative 0; Mismatches 176; Indels 3; Gaps 1;  
  
690 TCGGTTTGACGCTGCTGCTGCGCGGAATGAACCACTACAAACCTGATTTCAA 749  
|||||  
13 TCCAGATTGCTGCTGCTGCTGCGCGGAACCTACCTACCTGCTCACTATGAGCCCT 72  
|||||  
750 TCGGATGTACGATATTAGAAACCGCAGGGGTTTTCAGAGGAACCTGCGGAGCATCTG 809  
|||||  
73 GGGCTGCAACTCTGGCTCAGCACCTGATCAGCTGGCGCACTCAAGGCCAACCCATC 132  
|||||  
810 ACTGATGCTCAGGAGTGGAGAGAGCTTGCGTT---TCAGAGCGCGGCGCCCGTTTT 866  
|||||  
133 GCTTGGCGCCAGTTTGTGTAGAGAGCTCTGTGCTTACCACACTGGGCTCTGCACTGCGCAT 192  
|||||  
867 GAGGCGCATTCGCAAGCGGGATACGAGATCGGGGGCACCTGATTAAAGAAAGGTGATAT 926  
|||||  
193 CAAGGCACTGCCAAGGAGGATGTCATGATCGGCAAGCTTCGTCGGTGCAGAGAGG 252  
|||||  
927 GGTTTTGGCGTTTGTGGCATCGCAAAATCGTGATGAAGCAAGTTTTCAGACCCGCNAT 986  
|||||  
253 CATTATGCAATCCCAACCACTCAGGCAACAGGAGCAAGAAAGTCTTCGAGAAATCTGTATGA 312  
|||||  
987 GTTTCATATCGCGCGCATGCCAATCCCATATTCGCTTTGGCCACGGCAT 1037  
|||||  
313 GTTCAACATGAACCGCAAGTGGCCTACTCAAGATCCTCTTGGCTTTGGCTT 363  
|||||

**RESULT 7**

BZ551987/c 1223 bp DNA linear GSS 17-DEC-2002  
LOCUS pacs1-60.3425.x1 pacs1-60 Pseudomonas aeruginosa genomic clone  
DEFINITION pacs1-60.3425, genomic survey sequence.  
ACCESSION BZ551987  
VERSION BZ551987.1 GI:27155901  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 1223)

## AUTHORS

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.

## TITLE

Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library

## JOURNAL

J. Bacteriol. (2002) In press

## COMMENT

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

## FEATURES

source

## FEATURES

source

1. .638

/organism="Gossypium hirsutum"

/mol\_type="mRNA"

/cultivar="DeltaEMERALD"

/db\_xref="taxon:3635"

/clone="CFUS5.4D03"

/tissue\_type="Fov infected root and hypocotyls tissues"

/dev\_stage="seedlings, pooled from 5 hours to 7 days post infection with Fov."

/lab\_host="Y1090(ZL)"

/clone\_lib="Cotton Root and Hypocotyl tissues infected with *Fusarium oxysporum* f.sp. *vasinfectum* (Fov) Lambda ZIPLOX Library (CFUS)"

/note="Vector: Lambda ZIPLOX; Site 1: SalI; Site 2: NotI; mRNA was prepared from pooled root and hypocotyl tissues of the cotton cultivar DeltaEMERALD at different stages of Fov infection from 5 hours to 7 days post infection. cDNA was synthesised from a NotI-oligo primer/adaptor using the manufacturers protocols (Life Technologies) and then ligated to a SalI adaptor to facilitate directional cloning. The cDNA was cloned into the SalI and NotI sites of the Lambda ZIPLOX phage vector (Life Technologies). Constructed by Cairnion Dowd and Helen McFadden."

## ORIGIN

Query Match 4.4%; Score 52; DB 28; Length 1223;  
Best Local Similarity 50.8%; Pred. No. 0.00093;  
Matches 124; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 209 TSCGCGAGCAGACAGCTTATTTGGAATTCATTAAGATGACCGCGGAGCATA 268  
DB 534 TGAACGCTGGCGGCTTCTAAGACAACTTCATGCTTTCAACGAGCGCGCGGCACA 475  
QY 269 CAAAAATCCGTTGAGTCTGTAACAAAGCTTTACTCCGCGGCTGATGAAGCAATGGGAAC 328  
DB 474 CCGCGCTGGCAGGCTTTCGCTCGCGCTTCGCGCGCGCGCTGGGAG 415  
QY 329 CGAATTCAGAAATACAGATGAATCAATTTCAAAATTTACGGGCGCAGTTTG 388  
DB 414 CGCGCATCGAGGCTGCTGTAAGAACTCTCTCGACAGCTTCTGAGCGCGCGGAGCCG 355  
QY 389 ACCTGTTACAGATTTTATACCGCTTCGGTATTTGATATCTGAGCTGCGGAG 448  
DB 354 ATCTGCTCAGGATTTTCGCGAACCCTGACGATTCGGGTGGCGCGAGCTGTCGCT 295  
QY 449 TGCC 452  
DB 294 TTCC 291

## RESULT 8

CD486072

LOCUS

DEFINITION

CD486072 638 bp mRNA linear EST 01-JUL-2003  
CFUS5.4D09 Cotton Root and Hypocotyl tissues infected with *Fusarium oxysporum* f.sp. *vasinfectum* (Fov) Lambda ZIPLOX Library (CFUS)  
Gossypium hirsutum cDNA clone CFUS5.4D09 5' similar to Chain A, X-Ray Structure Of Nitric Oxide Reductase, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Gossypium hirsutum (upland cotton)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium. 1 (bases 1 to 638)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Dowd, C., Wilson, I. and McFadden, H.  
Different Gene Expression Responses in Cotton Root and Hypocotyl tissues during infection with *Fusarium wilt* D: Disease Unpublished (2003)  
Contact: Cairnion Dowd, Helen McFadden  
Commonwealth Scientific and Industrial Research Organisation  
Division of Plant Industry

Black Mountain Laboratories, Cnr Clunies Ross Street & Barry Drive, Black Mountain, Canberra, ACT, 2601, Australia  
Tel: 61 2 6246 4914, 6246 5377  
Fax: 61 2 6246 5000

Email: Cairnion.Dowd@csiro.au, Helen.McFadden@csiro.au  
vector clipped sequences Bases 1-17 (GTGACCCACGGCTCG): SalI adaptor

Seq primer: M13 reverse primer

High quality sequence stop: 638.

Location/Qualifiers

1. .638

/organism="Gossypium hirsutum"

/mol\_type="mRNA"

/cultivar="DeltaEMERALD"

/db\_xref="taxon:3635"

/clone="CFUS5.4D03"

/tissue\_type="Fov infected root and hypocotyls tissues"

/dev\_stage="seedlings, pooled from 5 hours to 7 days post infection with Fov."

/lab\_host="Y1090(ZL)"

/clone\_lib="Cotton Root and Hypocotyl tissues infected with *Fusarium oxysporum* f.sp. *vasinfectum* (Fov) Lambda ZIPLOX Library (CFUS)"

/note="Vector: Lambda ZIPLOX; Site 1: SalI; Site 2: NotI; mRNA was prepared from pooled root and hypocotyl tissues of the cotton cultivar DeltaEMERALD at different stages of Fov infection from 5 hours to 7 days post infection. cDNA was synthesised from a NotI-oligo primer/adaptor using the manufacturers protocols (Life Technologies) and then ligated to a SalI adaptor to facilitate directional cloning. The cDNA was cloned into the SalI and NotI sites of the Lambda ZIPLOX phage vector (Life Technologies). Constructed by Cairnion Dowd and Helen McFadden."

## ORIGIN

Query Match 4.1%; Score 49.4; DB 14; Length 638;  
Best Local Similarity 46.4%; Pred. No. 0.0042;  
Matches 198; Conservative 0; Mismatches 226; Indels 3; Gaps 1;

QY 614 AACCGAACAGGATATTTCTATTTTGTGAGCGGAAGAACAGCGGAGAGCTGT 673  
DB 91 AACGAGGATGATATCATAGTAACTCTGCACAGCAAGTCAAGCTGGAATATCG 150  
QY 674 CCGGTGAAGAGCTGATTCGGTTTTCACGCTCTCTGCTGGCGGAAATGAACCACTA 733  
DB 151 ACAAGTCGAGCGCTGTCCAGATTGCTTCTGCTCTGCTGCGGCAACGCTACATGG 210  
QY 734 CAACCTGATTTCAATGCGGATGTACAGCATATTAGAAACGCCAGCGCTTTACGAGAAC 793  
DB 211 TCAACATGATAGCCCTGGGGGTGCGAACTCTGGCTCAGCACCCCTGATCAGCTGGCGCAAC 270  
QY 794 TCGCAGCCATCTCGAACTGATGCTCAGGAGTGGAGAGAGCTTGCGTT---TCAGAG 850  
DB 271 TCAGGCCAACCCATCGTTGCGCCCGAGTTTGTGAGAGCTCTGTCTACCACTG 330  
QY 851 CGCGGCGCGCGTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGGGGGACCTGA 910  
DB 331 CGTCTGCACTCGCCATCAAGCGCATTCGCAAGGAGGATGTCTATGATCGGCGCAAGCTCG 390  
QY 911 TTAAGAAGGTGATATGGTTTGGCGTTTGTGGCATCGGCAAAATCGTATGAAGAAAGT 970  
DB 391 TCCGTGCAAAAGGAGGATATTCGATCCCAACAGTCCAGGACGAGGACGAGAAAGTCT 450  
QY 971 TTGACAGACCGCATATGTTGATATCGCGCCCATCCCAATCCGATATTCGTTTGGCC 1030  
DB 451 TCGAGATCTCTGATGAGTTCAACATGAACCGCAAGTGGCCCTACTCAAGATCCTCTTGCT 510  
QY 1031 ACGCAT 1037  
DB 511 TTGGCT 517

## RESULT 9



## CNS00396/c

LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921.1 GI:4941778

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP>. Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BAC08K10"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

## ORIGIN

Query Match 3.7%; Score 44.2; DB 29; Length 1101;  
Best Local Similarity 18.2%; Pred. No. 0.2;  
Matches 45; Conservative 117; Mismatches 85; Indels 0; Gaps 0;

QY 512 ATAAAGTGAAGCTGAAAGCTTTTGGAGACGAGATAGTGTGAGGAGAC 571

Db 848 ADDWADDWDWAWKWDWAWGARTDORRDWGDGAGKGGARKRRDRKRDAD 789

QY 572 TGGCCGCTTTTTCGGCATCATAGAGAAAGCGAAACAAACCGGACGATATTA 631

Db 788 DRDDAATWTWTTTTTRDTRDDWKTDTWTRWADRTWDRDDDDDRAGTKWRWTW 729

QY 632 TTTCTATTTAGTGAAGCGGAGAAACGCGAGAGCTGTCCGCTGAAGAGCTGATTC 691

Db 728 KRWKRRDRDDADDTARDORRRGGDAGKGGKTKRRRRDRATWDRTDWAWA 669

QY 692 CGTTTTCAGCTGCTGCTGGTGGCGGAAAGTGAACCACTACAACTGATTTCAAATG 751

Db 668 DAATTTTDDTDDWDRRRRRKRRRTTARAAWDDWTTWAKMDWAKWTRADWD 609

QY 752 CGATGTA 758

Db 608 RWAADTW 602

## RESULT 10

AZ934336/c

LOCUS

DEFINITION

BJ\_Ba0002013r B. japonicum BAC library Bradyrhizobium japonicum

588 bp DNA linear GSS 24-APR-2001

1136 bp DNA linear GSS 17-DEC-2002

pac2-164\_4486.y2 pac2-164 Pseudomonas aeruginosa genomic clone

pac2-164\_4486, genomic survey sequence.

genomic, genomic survey sequence.

AZ934336

VERSION

KEYWORDS

SOURCE

ORGANISM

Bradyrhizobium japonicum

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

Contact: Wing RA

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: [twing@clemson.edu](mailto:twing@clemson.edu)

Class: BAC ends

High quality sequence stop: 534.

Location/Qualifiers

1..588

/organism="Bradyrhizobium japonicum"

/mol\_type="genomic DNA"

/strain="USD110"

/db\_xref="taxon:375"

/lab\_host="E. coli"

/clone\_lib="B. japonicum BAC library"

/note="Vector: pIndigo536; Site\_1: HindIII"

## ORIGIN

Query Match 3.6%; Score 43.4; DB 28; Length 588;  
Best Local Similarity 45.8%; Pred. No. 0.26;  
Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 656 AACAGCGGAGAGCTGTCGGTGAAGAGCTGATTCGGTTTTCGACGCTCTCTGTGG 715

Db 328 AGAACGGCCAGCGGATGACGAGAGGAATTTGTGGCTACTGATCATCAGCTCAATA 269

QY 716 CCGGAATGAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGAAACGC 775

Db 268 CCGGACAGGTCCACACAGCTGATAGGACGGCGCTCGCTCGACCGTCATC 209

QY 776 CAGGCGTTTACGAGGAACCTGGCAGCCATCCTGAACTGATGCTTCAGGAGTGGAGAG 835

Db 208 CCGATCAGAAGCACCGGCTGCTCGGCCACCCAGACATGATCAAGACGGCGTGGAGTAGA 149

QY 836 CTTGCTGTTTACAGAGCGCGCGCCCGGTTTTCGAGGCGCATTCGAGCGGATACCGAGA 895

Db 148 TCGTCGCTACAGAGTTCACACAGCTGGGCAACCGCATGACCCGACCGGCTCGAGC 89

QY 896 TCGGGGGGACCTGATTAAGAAGGTGATATGCTTTTGGCGTTTGGCATCGGCAATC 955

Db 88 TCGGGCGGCTCATGCTCGATGCGGCGACCTCGGTGAGCGCTGTCATCGGGGCGGCCAAC 29

QY 956 GTGATGAAGCAAGTTTGACAGACC 980

Db 28 GCGACCCCGCGAGTTCCTCCGACCC 4

## RESULT 11

BZ563995/c

LOCUS

DEFINITION

pac2-164\_4486.y2 pac2-164 Pseudomonas aeruginosa genomic clone

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

BZ563995

pac2-164\_4486.y2

pac2-164\_4486, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

BZ563995.1 GI:2718931

GSS.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Pseudomonas aeruginosa	Pseudomonas aeruginosa	1 (bases 1 to 1136)	Spencer, D.H., Raymond, C.K., Smith, B.E., Sims, E.E., Hastings, M., Burns, J.F., Kaul, R. and Olsen, M.V.	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library	J. Bacteriol. (2002) In press	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: shotgun.
FEATURES	source	location/Qualifiers				
		1..1136				
		/organism="Pseudomonas aeruginosa"				
		/mol_type="genomic DNA"				
		/strains="2-164"				
		/db_xref="taxon:287"				
		/clone_lib="pacs2-164_4486"				
		/clone_lib="pacs2-164"				
		/note="clinical isolate 2-164 Whole genomic shotgun library."				
ORIGIN						
Query Match		3.6%; Score 42.8; DB 28; Length 1136;				
Best Local Similarity		48.7%; Pred. No. 0.53;				
Matches 113; Conservative		C; Mismatches 119; Indels 0; Gaps 0;				
QY		239 CCATCATTAACATGACCGCGGAGCATACAAAATCCGTTCACTCGTGAACAAGCCT 298				
Db		359 CCATGCTNCAGCGCCGCGCCGACCCAGGACACAAATGCGCCGCGTGTGCGAGCGCGT 300				
QY		299 TTACTCCGCGGTGATGAAGCAATGGGAACCGAGAATTCAGAAATCAAGATCAAGATCAACTGA 358				
Db		299 TCTCCCGAGCCACACAGAGCGACTCCAGACGATATAGANCGGTCACGACGAATGT 240				
QY		359 TTCAAAATTTACGGGGCGAGTGAAGTTGACCTTGTTCAGATTTTCATACCCGCTTC 418				
Db		239 TGGAGCCATGTCGCGCGCGAGAACACACCGACGTAATGGCGCGCTTCGATCCCACTGA 180				
QY		419 CGGTATATTGTATCTGAGCTGCTGGAGTGCCCTTCAGCGCATGGAACA 470				
Db		179 CCATCGCGGTGATCTAGACCTGCTGGGCATTCGAGGCGGAGCGCGGAACA 128				
RESULT 12						
BX381961/c						
LOCUS		1201 bp mRNA linear EST 08-MAY-2003				
DEFINITION		BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CS0DI0727F05 3-PRIME, mRNA sequence.				
ACCESSION		BX381961				
VERSION		BX381961.1				
KEYWORDS		EST.				
SOURCE		Homo sapiens (humar)				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 1201)				
TITLE		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
JOURNAL		Full-length cDNA libraries and normalization				
COMMENT		Unpublished (2001)				
		Contact: Genoscope				
		Genoscope - Centre National de Sequencage				
		BP 191 91006 EVRY cedex - France				
		Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr				
		Library was constructed by Life Technologies, a division of				
		Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :				

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODI072CC03NP1.  

FEATURES

Location/Qualifiers

source

1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref=taxon:9606"

/clone="CSODI072YF05"

/tissue type="PLACENTA COT 25-NORMALIZED"

/clone lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match                3.6%; Score 42.8; DB 3; Length 1201;  
Best Local Similarity     5.7%; Pred.No. 0.54;  
Matches      44; Conservative 253; Mismatches 469; Indels    3; Gaps    1

Qy	370	CAGGGCGCAGTAGTTGACTTGTCACAGATTTTCATACCCTGCCTCCGTATTATGTA	TGTTG	429
Dd	1088	CGKKGCCSKGVVKMGEMMKKKMMKMVMGMOMMKBKOMGKKKKNNKKKBBKTU		1029
Qy	430	ATACTCAGCTGCTGGAGTGCCTTCACGCCAGATGAACAAGTTAAAGCATGCTCTGAT	A	489
Dd	1028	KKKTVTKBMTGTRKKKKXGKHMGGEVMKVMMVGKGNVHGKGBMBBGKRQMNTGK		969
Qy	490	CTTCGTCTAGTACACCGAAGGATAAGAAGCTGAAGAACCTTTTTTGGAGAAA		549
Dd	968	MVKXBMMGGSGOHHMKBEIVNKTTNMVBKKMNKKMMNGKVVWVKKNMMDMM		909
Qy	550	CGAGATAAGTGTGAGGAAGAAGCTGCCGCGCTTTTTTCOGCGCATCATAGAAGAAAGCA	G	609
Dd	908	MKIGRWAKKGNMVVKONNKKKKKKKTKTKKKKKKKHMKKTMNMKKCKOKKKKKCN		849
Qy	610	AAC---AAACCGGAAACAGGATATATTCTCTATTTAGTGAAGCGGAAGAAAAAGCGGAG		666
Dd	848	MMYGKRMKMGKXXKYMKGMMNMKKCKGEBKMCVMKMMMMKGGJMHGGRMMMGMS		789
Qy	667	AACTGTCGGTGAAGAGCTGATTCGGTTTTTGACCGCTGCTGCTGGTGGCGCGGAAATGAA		726
Dd	788	KMKOMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGNNKKNNNNNNNNKK		729
Qy	727	ACCACCTACAAAACCTGATTTCAAATG-CAGTGACAGCATATTAGAAAGCACGGCGTTTAC		786
Dd	728	NN		669
Qy	787	GAGGAACCTCGCGAGCCACTCTGAAC TGCTGATGCGCTCAGGCGAGTGGAGGAAGCTTCGCTTC		846
Dd	668	NN		609
Qy	847	AGAGCGCCGCCCGGTTTGGAGCGCATGTGCAAGSCGGGATACGGAGATCGGGGGGCGAC		906
Dd	608	NN		549
Qy	907	CTGATTAAAGAAGETGATAGCTTTTGGCGTTTGTGGCATCGCAATCGTGATGAAGCA		966
Dd	548	XKKMMNNNNKTNKKONGNNNNKKKKKKKKKKKKKKKKKKKKKKNNNNNNNNNNNNNN		489
Qy	967	AAGITTTGACAGACCGCATGTTTGATATCCCGCGCATCCCAATCCCGCATATTCGTTT		1026
Dd	488	NMMNNNNKNGNN		429
Qy	1027	GGCCAAGCATCCATTTTGGCTTGGGCCCGCGCTTGCCCGCTTTGAAGCAAAATATCGCG		1086
Dd	428	KKKMMNNKGMAMMMNNNNKKMMNNNNNTNGTMMOMNNNNTWGINNVGGWTTKSGENGNN		369
Qy	1087	TTAACGCTCTTGATTCTCGCTTTTCTCATZATGGAGTGGTCAGTATCA		1135
Dd	368	GTMMNNNTTTTMMNTNCNNTTCANTTK-TTITNTNTTCNNNSSTTCCA		320

## RESULT 13

CG092084  
LOCUS  
DEFINITION  
CG092084 ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBtra0545F03,  
genomic survey sequence.  
ACCESSION  
CG092084  
VERSION  
CG092084.1 GI:33974378  
KEYWORDS  
GSS.  
SOURCE  
Zea mays  
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 718)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.

## TITLE

Maize Genomics Consortium

## JOURNAL

Unpublished (2003)

## COMMENT

Other GSSs: PUTBM26TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..718

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone\_lib="ZMMBtra0545F03"

/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

Query Match 3.5%; Score 42.2; DB 29; Length 718;

Best Local Similarity 49.3%; Pred. No. 0.65;

Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 866 TGAGCGGCGATTCCCAAGCGGGATACGAGATCGGGGCGACCTGATTAAGAGGTGATA 925

Db 13 TGCGCGGAGCGGATGAGGACACATCGTTGGCGGTGAGCCATCGCGAGGGGGGACA 72

QY 926 TGGTTTGGCGTTTGCGATCGGCAATCGTGATGAAGCAAAAGTTTGACAGACCGCAC 985

Db 73 AGGTGCTGTGTGTATCATCTCCGCCAATCGCGAGAAAGCGCTTCCCGACGCGATC 132

QY 986 TGTGTGATCCGCGCATCCCATCGCATTTGCGTTTGGCCACGCGCATCCATTTT 1045

Db 133 GCTTCGAGCTGTCGCAAGGCGCGGACATGTCGGCTTCGGGCGCGGACGCTCT 192

QY 1046 GCCTTGGGCGCGCGCTTGGCGCTTCTTGAAGCAAAATATCGGTT 1088

Db 193 GCGTCGATCGCGCTCGCGGAATGCGTATCGGCTT 235

Query Match 3.4%; Score 40.6; DB 13; Length 432;

Best Local Similarity 49.3%; Pred. No. 1.6;

Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 562 GAGGAAGAACTGGCGCGTCTTTTTCGCGCATCATAGAGAAAGCGAAACCGGAA 621

Db 135 GTGGCAGAGCGGACTAGCGTTGCTATGAGGAGAGAGAGATGCTGTAACAA 194

QY 622 CAGGATATTTCTATTTTGTGAAGCGGAGAAACAGGCGAGAGCTGTCGGTGA 681

Db 195 GAGAAATGAGTGACATGCTTGGGGCACTGTGGCGCTCGCGGACCACTTCCGACG 254

QY 682 GAGCTGATCTCGGTTTTCACGCTGCTGCTGGCGGGAATGAACCACTCAACCTG 741

Query Match 3.4%; Score 40.6; DB 13; Length 432;

Best Local Similarity 49.3%; Pred. No. 1.6;

Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 562 GAGGAAGAACTGGCGCGTCTTTTTCGCGCATCATAGAGAAAGCGAAACCGGAA 621

Db 135 GTGGCAGAGCGGACTAGCGTTGCTATGAGGAGAGAGATGCTGTAACAA 194

QY 622 CAGGATATTTCTATTTTGTGAAGCGGAGAAACAGGCGAGAGCTGTCGGTGA 681

Db 195 GAGAAATGAGTGACATGCTTGGGGCACTGTGGCGCTCGCGGACCACTTCCGACG 254

QY 682 GAGCTGATCTCGGTTTTCACGCTGCTGCTGGCGGGAATGAACCACTCAACCTG 741

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

## REFERENCE

## AUTHORS

1 (bases 1 to 432)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,  
Khanna, A., Holla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,  
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,  
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,  
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
McCann, R., Waterston, R. and Wilson, R.

## TITLE

## JOURNAL

## COMMENT

Public Soybean EST Project

Contact: Shoemaker R/Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact: ccu@resgen.com web site:

www.resgen.com

Seq primer: -40RP from Gibco.

Location/Qualifiers

1..432

/organism="Glycine max"

/mol\_type="mRNA"

/db\_xref="taxon:3847"

/clone="SOYBEAN CLONE ID: Gm-cl074-8560"

/tissue\_type="seedlings induced for HR (hypersensitive

response)"

/dev\_stage="9-11 day old"

/lab\_host="DH10B"

/clone\_lib="Gm-cl074"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from mRNA isolated

from 9-11 day old seedlings that were induced for HR

(hypersensitive response) by vacuum infiltrating plant

tissue with Pseudomonas syringae pv. glycinea carrying the

avrB gene (Genetics 141:1597-1604). Plant tissue (expanded

unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,

and 53 hrs after inoculation and their mRNA pooled equally

for cDNA construction. The library was prepared using the

Stratagene pBluescript II SK(+) library construction kit.

Complementary DNA was synthesized from mRNA using a primer

consisting of a poly(dT) sequence with an XhoI restriction

site. EcoRI adaptors were ligated to the blunt-ended cDNA

fragments followed by XhoI digestion. The cDNA insert is

protected from XhoI digestion via methylation during first

strand synthesis. The cDNA fragments were directionally

cloned into the EcoRI-XhoI restriction site of the

pBluescript vector. The ligated cDNA fragments were

transformed into E. coli ElectroMax DH10B host cells. Plant

care, inoculations, and library construction were

performed by Steve Clough (Lila Vodkin lab, University of

Illinois)."

Query Match 3.4%; Score 40.6; DB 13; Length 432;

Best Local Similarity 49.3%; Pred. No. 1.6;

Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 562 GAGGAAGAACTGGCGCGTCTTTTTCGCGCATCATAGAGAAAGCGAAACCGGAA 621

Db 135 GTGGCAGAGCGGACTAGCGTTGCTATGAGGAGAGAGATGCTGTAACAA 194

QY 622 CAGGATATTTCTATTTTGTGAAGCGGAGAAACAGGCGAGAGCTGTCGGTGA 681

Db 195 GAGAAATGAGTGACATGCTTGGGGCACTGTGGCGCTCGCGGACCACTTCCGACG 254

QY 682 GAGCTGATCTCGGTTTTCACGCTGCTGCTGGCGGGAATGAACCACTCAACCTG 741

```

Db      255 GAAATAGTGGATTTTTTCGCTTTCGCTGCGCGGTAGGAAACACCTCTACCATA 314
Qy      742 ATTTCAATGCGATGATCAGCATATTAAGACGCC 776
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      315 ATGACTCTTGGATCAAGTTCCTCACTGAGACTCC 349

RESULT 15
BX406481
LOCUS   BX406481 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION
CSODM012YA15 5-PRIME, mRNA sequence.
ACCESSION   BX406481
VERSION     BX406481.1 GI:30635544
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
AUTHORS     Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            Bp 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSODM012AA08QF1.
            Location/Qualifiers
            1..1201
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CSODM012YA15"
               /tissue_type="FETAL LIVER"
               /dev_stage="fetal"
               /clone_lib="Homo sapiens FETAL LIVER"
               /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
               was primed with a NotI-oligo(dT) primer. Five prime end
               enriched, double-strand cDNA was digested with Not I and
               cloned into the Not I and EcoRV sites of the pCMVSPORT 6
               vector. Library was not normalized."

FEATURES
         source
         1..1201

ORIGIN

Query Match      3.4%; Score 40.6; DB 13; Length 1201;
Best Local Similarity 23.8%; Pred. No. 2.5;
Matches 54; Conservative 74; Mismatches 99; Indels 0; Gaps 0;

Qy      422 TTATGTGATATCTGAGCTGCTGGAGTGCCCTTCAGCGGATGGAACAGTTTAAAGCAT 481
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      573 KKKKTCNNKKKKKKKNNCTTKATNKTKNKKTKNKKCCCNKKKKAADAKKAAKAKKDC 632
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      482 GGTCTGATCTTCTGCTAGTACACCGAAGGATAAAGTGAAGAGCTGAATAAGCCTTT 541
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      633 CKADKKKKKKKKKKKKMAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 692
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      542 TGAAGACGAGATAAGTGTGAGGAGAACTGGCGCGGTTTTTTTTCGCGCATCATAGAAG 601
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      693 AKAATAKAKKAKKAAADKKKKTKAATAAATAAATAAATAAATAAATAAATAAATAA 752
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      602 AAAAGCGAACAACACCGGACAGGATATTATTCTTATTATTAGTGAA 648
         ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db      753 AATAKKAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 799
         ||||| : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 28, 2004, 22:08:31  
Job time : 2413 secs